Query Match 3.5%; Score 31.4; DB 4; Length 1119;
Best Local Similarity 61.7%; Pred. No. 3.1;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 28 ccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaatgcctctccctg 87
|| || |||| | ||||| || | ||| || || ||||| | | | | ||
Db 187 ccggctatgcgctcggccatcaacctgctgctggccaccctggccttctccgacatcatg 246

Qy 88 ctgccgctttcctgtctgcac 108
||| | || | ||| ||| |
Db 247 ctgtccctctgctgcatgccc 267

RESULT 18
US-08-987-943-1
; Sequence 1, Application US/08987943
; Patent No. 6210967
; GENERAL INFORMATION:
; APPLICANT: Bard, Jonathan A.
; TITLE OF INVENTION: DNA Encoding A Mammalian LPA Receptor and Uses Thereof
; FILE REFERENCE: 55183
; CURRENT APPLICATION NUMBER: US/08/987,943
; CURRENT FILING DATE: 1997-12-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 beta
; SEQ ID NO 1
; LENGTH: 1755
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-987-943-1

Query Match 3.5%; Score 31.4; DB 4; Length 1755;
Best Local Similarity 61.7%; Pred. No. 4;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 28 ccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaatgcctctccctg 87
|| || |||| | ||||| || | ||| || || ||||| | | | | ||
Db 536 ccggctatgcgctcggccatcaacctgctgctggccaccctggccttctccgacatcatg 595

Qy 88 ctgccgctttcctgtctgcac 108
||| | || | ||| ||| |
Db 596 ctgtccctctgctgcatgccc 616

RESULT 19
[redacted]
; Sequence 15, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5247 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical Isolate P2-17
US-08-920-812-15

Query Match 3.5%; Score 31.4; DB 1; Length 5247;
Best Local Similarity 64.4%; Pred. No. 7.2;
Matches 47; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 42 cgccatgcacatcctgttgaccgccctgctcaaatgcctctccctgctgccgctttcctg 101
||||||| || || | [redacted] | || |||| | |||| | |
Db 698 CGCCATGTACTTCGGCTG[redacted]CCTGCTCATGACCATCCTCCTGGTCGCGCTGTTCCT 639

Qy 102 tctgcacacgctg 114
|||||||
Db 638 CGGCGCCACGCTG 626

RESULT 20
US-08-920-827-15/c
; Sequence 15, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448

```
;     REFERENCE/DOCKET NUMBER:  17796-002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 877-0900
;      TELEFAX:  (415) 877-8370
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2070 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US91-07290-6


  Query Match             3.7%;  Score 33.4;  DB 5;  Length 2070;
  Best Local Similarity   47.4%;  Pred. No. 2.9;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     381 catcggcagctacgatttgggcggacgctacatcagccagcagcttccgttccacctgac 440
           || | || | | | |   ||      ||   ||| | || ||||  |   |  | | |
Db     732 CAACTGCCGGTGCAAGGAGGAGAACTGCGCTATCCTCAAGGAGCTGGTGACGCTGCGGGC 791

Qy     441 cgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggt 500
           |   | || |   | ||     || |  ||  | | |     | | || |   |   |
Db     792 CCAGAAGTCCCGCCTGCTGGGGTTCCACACGCACGCCGACTATGTCCTGGAGATGAACAT 851

Qy     501 gcgcggcaaaggcaaaaccgcgcccaccggcatacaagggtcaaacaaatcatcaaggc 560
           |  |   |   || | |||| |||||  | || | | |  |    || |   | ||| |
Db     852 GGCCAAGACCAGCCAGACCGTGGCCACCTTCCTAGATGAGCTGGCGCAGAAGCTGAAGCC 911

Qy     561 cctgcgcgcgggcgaggcaaccatcatcctg 591
           |||| | | |   |||    |  | || |||
Db     912 CCTGGGGGAGCAGGAGCGTGCGGTGATTCTG 942


RESULT  18
US-09-422-487-4
; [redacted]Application US/09422487
; Patent No. [redacted]
; GENERAL INFORMATION:
;  APPLICANT:  Cardineau, Guy A.
;  APPLICANT:  Stelman, Steven J.
;  APPLICANT:  Narva, Kenneth E.
;  TITLE OF INVENTION:  Plant-Optimized Polynucleotides Encoding Approximately
;  TITLE OF INVENTION:  15 kDa and Approximately 45 kDa Pesticidal Proteins
;  FILE REFERENCE:  MA723/4X
;  CURRENT APPLICATION NUMBER:  US/09/422,487
;  CURRENT FILING DATE:  1999-10-21
;  PRIOR APPLICATION NUMBER:  60/105,408
;  PRIOR FILING DATE:  1998-10-23
;  PRIOR APPLICATION NUMBER:  60/105,359
;  PRIOR FILING DATE:  1998-10-23
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 1155
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
;   OTHER INFORMATION: Bacillus thuringiensis toxin gene
US-09-422-487-4


  Query Match             3.7%;  Score 33.2;  DB 4;  Length 1155;
  Best Local Similarity   46.8%;  Pred. No. 2.5;
  Matches  104;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy     324 ggaacacgtgcagcaggctttggacaagggcgaagggctgctgttcatcacgccgcacat 383
           | || ||  | |   |    | | |||||||       ||    || || ||| || |
Db     618 gaaatacaagtactggaacctcgccaagggctccaacgtgtccctcctcccgcaccagaa 677

Qy     384 cggcagctacgatttgggcggacgctacatcagccagcagcttccgttccacctgaccgc 443
           |||||| |  | | | | ||          ||  | | | | ||
Db     678 gcgcagctacgactacgagtgggacctgagaagaaccagaagaccaccatcatcaacac 737

Qy     444 catgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtgcg 503
           | ||  | || |   || ||| |  ||      ||  ||  |  | |||    ||   |
Db     738 cgtgggcctgcagatcaacatcgactcggggatgaagttcgaggtgccggaggtgggcgg 797

Qy     504 cggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaa 545
           |||||  |   | | |  | ||          ||  | | | ||
Db     798 cggcaccgaggacatcaagacccagctcaccgaggagctgaa 839


RESULT  19
US-08-844-188-45
; Sequence [redacted] Application US/08844188
; Patent No. [redacted]
;  GENERAL INFORMATION:
;    APPLICANT:  Narva, Kenneth E.
;    APPLICANT:  Schnepf, H. Ernest
;    APPLICANT:  Knuth, Mark
;    APPLICANT:  Pollard, Michael R.
;    APPLICANT:  Cardineau, Guy
;    APPLICANT:  Schwab, George E.
;    TITLE OF INVENTION:  Pesticidal Toxins
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/844,188
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/633,993
;      FILING DATE:  19-APR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sanders, Jay M.
;      REGISTRATION NUMBER:  39,355
;      REFERENCE/DOCKET NUMBER:  MA-703C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  352-375-8100
;      TELEFAX:  352-372-5800
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1158 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-844-188-45


  Query Match             3.7%;  Score 33.2;  DB 3;  Length 1158;
  Best Local Similarity   46.8%;  Pred. No. 2.5;
  Matches  104;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy     324 ggaacacgtgcagcaggctttggacaagggcgaagggctgctgttcatcacgccgcacat 383
           | || ||  | |   |    | | |||||||       ||    || || ||| || |
Db     618 GAAATACAAGTACTGGAACCTCGCCAAGGGCTCCAACGTGTCCCTCCTCCCGCACCAGAA 677

Qy     384 cggcagctacgatttgggcggacgctacatcagccagcagcttccgttccacctgaccgc 443
```

```
Db    678 GC[redacted]CTACGAGTGGGGCACCGAGAAGAACCAGAAGACCACCATCATCAACAC 737

Qy    444 catgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtgcg 503

Db    738 CGTGGGCCTGCAGATCAACATCGACTCGGGGATGAAGTTCGAGGTGCCGGAGGTGGGCGG 797

Qy    504 cggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaa 545

Db    798 CGGCACCGAGGACATCAAGACCCAGCTCACCGAGGAGCTGAA 839


RESULT 20
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
;   APPLICANT: Katz, L
;   APPLICANT: Donadio, S
;   APPLICANT: Mcalpine, J B
;   TITLE OF INVENTION: Recombinant DNA Method for Producing
;   TITLE OF INVENTION: Erythromycin Analogs
;   NUMBER OF SEQUENCES: 27
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Edward H. Gorman
;     STREET: Abbott Laboratories D377/AP6D-2 One Abbott
;     STREET: Park Rd
;     CITY: Abbott Park
;     STATE: IL
;     COUNTRY: US
;     ZIP: 60064-3500
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/642,734C
;     FILING DATE: 17-JAN-91
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Danckers, Andreas M
;     REGISTRATION NUMBER: 32652
;     REFERENCE/DOCKET NUMBER: 4952.US.01
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 708-937-9396
;     TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 20235 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: unknown
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;     ORGANISM: Saccharopolyspora erythraea
;     STRAIN: NRRL 238
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 19..10722
;     OTHER INFORMATION: /codon_start= 19
;     OTHER INFORMATION: /function= "gene eryA"
;     OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
;     OTHER INFORMATION: 6-deoxyerythronolide B"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 19..4470
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: module 3"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 97..1482
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 1693..2670
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: acyltransferase domain module 3"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 3406..3921
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: beta-ketoreductase domain of module 3"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 4171..4428
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: acyl carrier domain of module 3"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 4471..10722
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: module 4"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 4471..5847
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: beta-ketoacylACPsynthase domain of module"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 6054..7026
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: acyltransferase domain of module 4"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 7165..9216
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: dehydratase and enoylreductase domains m"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 9433..9984
;     OTHER INFORMATION: /function= "approximate span
;     OTHER INFORMATION: beta-ketoreductase of module 4"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 10225..10483
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: acyl carrier domain of module 4"
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 10723..20235
;     OTHER INFORMATION: /codon_start= 10723
;     OTHER INFORMATION: /function= "gene =eryA"
;     OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
;     OTHER INFORMATION: 6-deoxyerythronolide B formatio"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 10723..15165
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: module 5"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 10831..12174
;     OTHER INFORMATION: /function= "approximate span of
;     OTHER INFORMATION: beta-ketoacylACPsynthase domain of modul"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 12379..13350
;     OTHER INFORMATION: /function= "approximatr span of
;     OTHER INFORMATION: acyltransferase domain of module 5"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 14062..14610
```

```
Db  2502 GGCACAACAGTGACGTTGGAAAATGCGACTTGGACAATGCCTAGCGATACTACATTGCAG 2561
Qy  2704 aatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcggca 2763
Db  2562 AATTTAACGCTAAATAACAGTACGATCACGTTAAATTCAGCTTATT-------------- 2607
Qy  2764 ggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgccgt 2823
Db  2608 ----------CAGCTAGCTCAAACAATACGCCACGTCGCCGTTCA-------------- 2642
Qy  2824 tccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacggta 2883
Db  2643 ---TTAGAGACGGAAACAACGCCAACATCGGCAGAACATCGTTTCAACACATTGACAGTA 2699
Qy  2884 aacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctaccgc 2943
Db  2700 AATGGTAAATTGAGTGGGCAAGGCACATTCCAATTTACTTCATCTTTATTTGGCTATAAA 2759
Qy  2944 agcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaacaat 3003
Db  2760 AGCGATAAATTAAAATTATCCAATGACGCTGAGGGCGATTACATATTATCTGTTCGCAAC 2819
Qy  3004 accggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaacaaa 3063
Db  2820 ACAGGCAAAGAACCCGAAACCCTTGAGCAATTAACTTTGGTTGAAAGCAAAGATAATCAA 2879
Qy  3064 ccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcgtgg 3123
Db  2880 CCGTTATCAGATAAGCTCAAATTTACTTTAGAAAATGACCACGTTGATGCAGGTGCATTA 2939
Qy  3124 cgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaa 3183
Db  2940 CGTTATAAATTAGTGAAGAATGATGGCGAATTCCGCTTGCATAACCCAATAAAAGAGCAG 2999
Qy  3184 gagctttccgacaaactcggcaaggc----------------------agaagccaaaaaa 3222
Db  3000 GAATTGCACAATGATTTAGTAAGAGCAGAGCAAGCAGAACGAACATTAGAAGCCAAACAA 3059
Qy  3223 caggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgcc 3282
Db  3060 GTTGAACCGACTGCTAAAACACAAACAGGTGAGCCAAAAGTGCGGTCAAGAAGAGCAGCG 3119
Qy  3283 gtcgaaaag---acaggaaagcgttgccgaaccggcccggcaggcaggcggggaaaatgtc 3339
Db  3120 AGAGCAGCGTTTCCTGATACCCTGCCTGATCAAAGCCTGTTAAACGCATTAGAAGCCAAA 3179
Qy  3340 ggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaagacaccgccttg 3399
Db  3180 CAAGCTGAACTGACTGCTGAAACACAAAAAAGTAAGGCAAAAACAAAAAAAGTGCGGTCA 3239
Qy  3400 gcgaaacagcgcgaagcggaaacccggccggctaccaccgccttcccccgcgcccgccgc 3459
Db  3240 AAAAGAGCAGTGTTTTCTGATCCCCTGCTTGATCAAAGCCTGTTCGCATTAGAAGCCGCA 3299
Qy  3460 gcccgccgggatttgccgcaactgcaaccccaaccgcagccccaacc---------------- 3506
Db  3300 CTTGAGGTTATTGATGCCCCACAGCAATCGGAAAAAGATCGTCTAGCTCAAGAAGAAGCG 3359
Qy  3507 ----------gcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaa 3552
Db  3360 GAAAAACAACGCAAACAAAAAGACTTGATCAGCCGTTATTCAAATAGTGCGTTATCAGAA 3419
Qy  3553 ttttccgccacgctcaacagcgtttttcgccgtacaggacgaattagaccgcgtatttgcc 3612
Db  3420 TTATCTGCAACAGTAAATAGTATGCTTTCTGTTCAAGATGAATTAGATCGTCTTTTTGTA 3479
Qy  3613 gaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcg 3672
Db  3480 GATCAAGCACAATCTGCCGTGTGGACAAATATCGCACAGGATAAAAGACGCTATGATTCT 3539
Qy  3673 caagatttccgcgcctaccgccaac---aaaccgacctgcgccaaatcggtatgcagaaa 3729
Db  3540 GATGCGTTCCGTGCTTATCAGCAGCAGAAAACGAACTTACGTCAAATTGGGGTGCAAAAA 3599
```

```
Qy  3730 aacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaaaacaccttc 3789
Db  3600 GCCTTAGCTAATGGACGAATTGGGGCAGTTTTCTCGCATAGCCGTTCAGATAATACCTTT 3659
Qy  3790 gacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcgggcaatacggc 3849
Db  3660 GATGAACAGGTTAAAAATCACGCGACATTAACGATGATGTCGGGTTTTGCCCAATATCAA 3719
Qy  3850 atcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggcagcctttca 3909
Db  3720 TGGGGCGATTTACAATTTGGTGTAAACGTGGGAACGGGAATCAGTGCGAGTAAAATGGCT 3779
Qy  3910 gacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcaggcacgatac 3969
Db  3780 GAAGAACAAAGCCGAAAAATTCATCGAAAAGCGATAAATTATGGCGTGAATGCAAGTTAT 3839
Qy  3970 cgcgccggtttcggcggattcggcatcgaaccgccatcggcgcaacgcgctatttcgtc 4029
Db  3840 CAGTTCCGTTTAGGGCAATTGGGCATTCAGCCTTATTTTGGAGTTAATCGCTATTTTATT 3899
Qy  4030 caaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggccttgcattcaac 4089
Db  3900 GAACGTGAAAATTATCAATCTGAGGAAGTGAGAGTGAAAACGCCTAGCCTTGCATTTAAT 3959
Qy  4090 cgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatc 4149
Db  3960 CGCTATAATGCTGGCATTCGAGTTGATTATACATTTACTCCGACAGATAATATCAGCGTT 4019
Qy  4150 acgccttatttgagccgtcctataccgatgccgcttcgggcaaagtccgaacacgcgtc 4209
Db  4020 AAGCCTTATTTCTTCGTCAATTATGTTGATGTTTCAAACGCTAACGTACAAACCACGGTA 4079
Qy  4210 aatacgccgtattggctcaggatttcggcaaaacccgcagtgcggaatggggcgtaaac 4269
Db  4080 AATCTCACGGTGTTGCAACAACCATTTGGACGTTATTGGCAAAAAGAAGTGGCATTAAAG 4139
Qy  4270 gccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggcccgcaactg 4329
Db  4140 GCAGAAATTTTACATTTCCAAATTTCCGCTTTTATCTCAAAATCTCAAGGTTCACAACTC 4199
Qy  4330 gaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
Db  4200 GGCAAACAGCAAAATGTGGGCGTGAAATTGGGCTATCGTTGGTAA 4244
```

```
RESULT  15
NLIGA3327
LOCUS       NLIGA3327                578 bp    DNA     linear [redacted]
DEFINITION  Niesseria lactamica IgA protease homologue gene, isol[redacted]
ACCESSION   AJ001740
VERSION     AJ001740.1  GI:3152399
KEYWORDS    IgA protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 578)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Neisseria lactamica"
                     /isolate="NL3327"
                     /db_xref="taxon:486"
```

```
     CDS             <1. .>576
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA04965.1"
                     /db_xref="GI:3152400"
                     /db_xref="SPTREMBL:O69746"
                     /translation="GKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSYVSRNGVAALV
                     GDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRL
                     HKFVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGAWLI
                     GGNTHMQGWGNNGVVSLSGDVRHANDYGPMPI"
     repeat_region   88. .95
                     /rpt_type=INVERTED
     repeat_region   139. .148
BASE COUNT      163 a    125 c    156 g    134 t
ORIGIN
```

```
  Query Match             7.3%;  Score 318.8;  DB 1;  Length 578;
  Best Local Similarity   83.3%;  Pred. No. 4.1e-61;
  Matches  375;  Conservative    0;  Mismatches   72;  Indels    3;  Gaps    1;

Qy    184 ggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtc 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 60

Qy    244 ggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtg 303
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     61 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTG 120

Qy    304 gcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaac 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAAT 180

Qy    364 gttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaattgtg 423
          ||||||||||||||| || ||||| |||||||||||| || || ||| ||||  ||||||||
Db    181 GTTGATTTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG 240

Qy    424 aaacggaataattataaagcagggactaaaggccatcctatggcggcgattatcatatg 483
          ||| | ||||| || ||| | |  | |  |   || |||||  |||||||||| ||||||
Db    241 AAAAGAAATAACTACAAACCTGACAATTCA---CACCCTTACAACGGCGATTACCATATG 297

Qy    484 ccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatggat 543
          |||||  ||||||| || || || |||||||||||||| ||||||||  || | |||
Db    298 CCGCGCCTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGG 357

Qy    544 gggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggcagg 603
          |||   |  |||  |||| || | || || || ||||| || || ||  ||||
Db    358 GGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACAC 417

Qy    604 caatattggcgatctgatgaagatgagccc 633
          || |||||||| | |||||| ||  | | |
Db    418 CACTATTGGCGTTATGATGATGACAAACAC 447
```

```
RESULT 16
NLIGA3354
LOCUS       NLIGA3354                584 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3354.
ACCESSION   AJ001739
VERSION     AJ001739.1  GI:3152397
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 584)
```

```
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Neisseria lactamica"
                     /isolate="NL3354"
                     /db_xref="taxon:486"
     CDS             <1. .>582
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04964.1"
                     /db_xref="GI:3152398"
                     /db_xref="SPTREMBL:O69747"
                     /translation="FAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGD
                     QYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLHK
                     FVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDDKHGDLSYSGAWLIG
                     GNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGA"
     repeat_region   82. .88
                     /rpt_type=INVERTED
     repeat_region   134. .142
BASE COUNT      163 a    127 c    159 g    135 t
ORIGIN
```

```
  Query Match             7.2%;  Score 313.6;  DB 1;  Length 584;
  Best Local Similarity   83.6%;  Pred. No. 6e-60;
  Matches  368;  Conservative    0;  Mismatches   69;  Indels    3;  Gaps    1;

Qy    190 tttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaa 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAA 60

Qy    250 tcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggca 309
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     61 TCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCA 120

Qy    310 ttggtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgat 369
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    121 TTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGAT 180

Qy    370 tttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaattgtgaaacgg 429
          |||||||| || ||||| |||||||||||| || || ||| ||||  |||||||||| |
Db    181 TTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGA 240

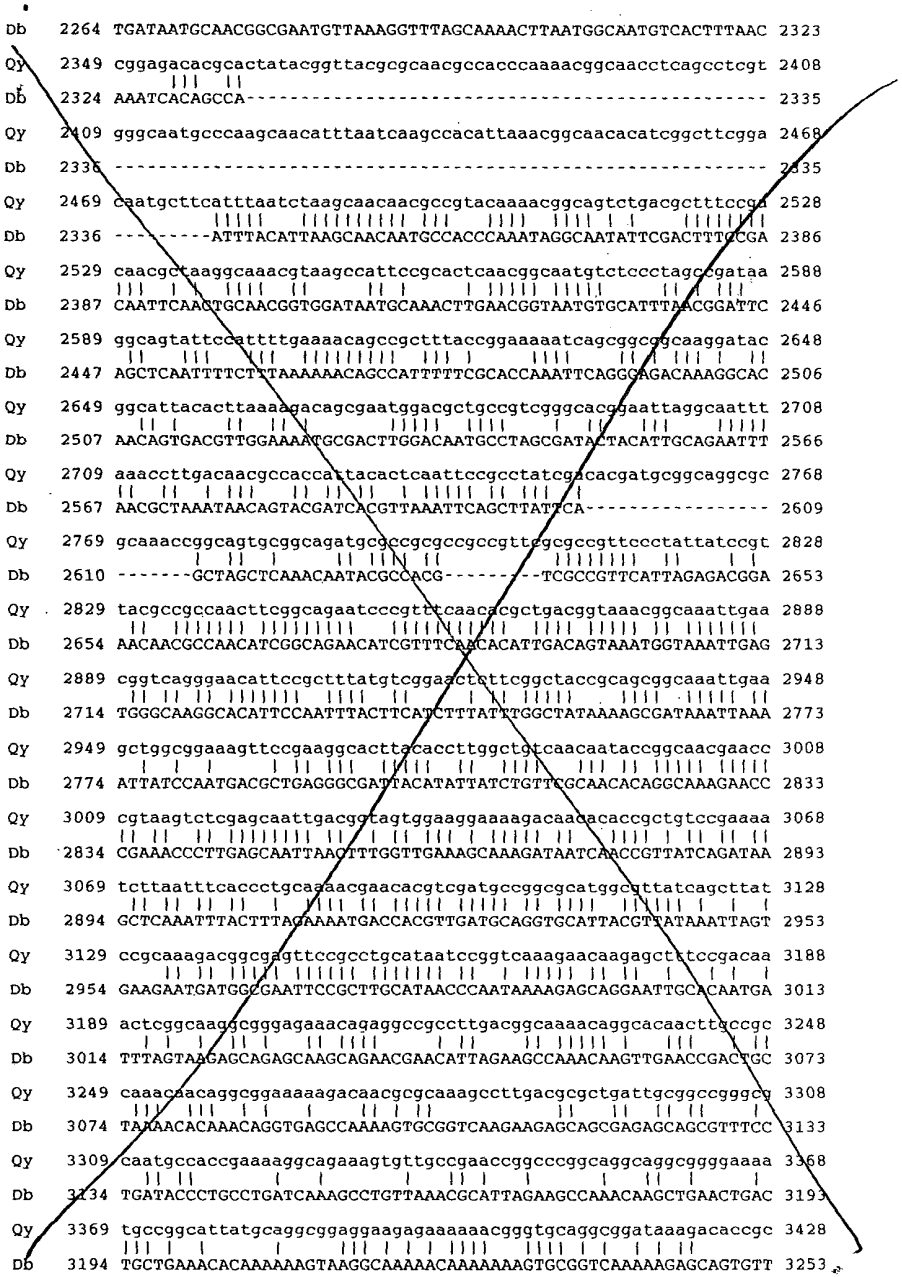
Qy    430 aataattataaagcagggactaaaggccatcctatggcggcgattatcatatgccgcgt 489
          ||||| || || | |  | |  |   || ||||||  ||||||||| ||||||||||
Db    241 AATAACTACAAACCTGACAATTCA---CACCCTTACAACGGCGATTACCATATGCCGCGC 297

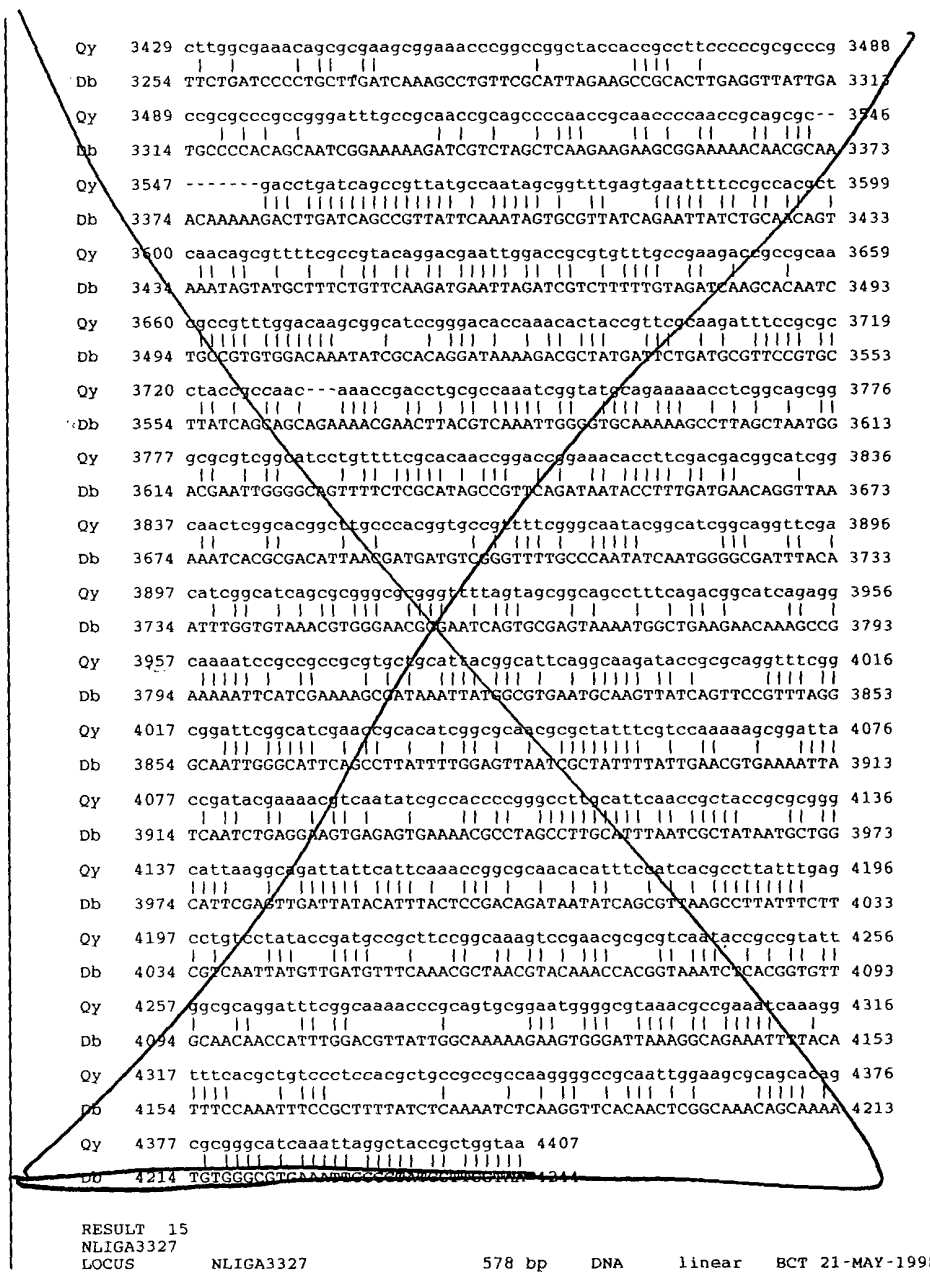
Qy    490 ttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatggatgggcgg 549
          |||||| || || || ||||||||||||| |||||||  || | |||   |||
Db    298 CTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGGAAT 357

Qy    550 aaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggcaggcaatat 609
          |  |||  |||| || | || || || || ||||| || || ||  ||||    || |||
Db    358 ACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTAT 417

Qy    610 tggcgatctgatgaagatga 629
          ||||| | |||||| || ||
Db    418 TGGCGTTATGATGATGACGA 437
```

```
RESULT 17
NLIGA823
LOCUS       NLIGA823                 572 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL823.
ACCESSION   AJ001737
VERSION     AJ001737.1  GI:3152389
KEYWORDS    IgA protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
```





RESULT 15
NLIGA3327
LOCUS NLIGA3327 578 bp DNA linear BCT 21-MAY-1998

```
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3327.
ACCESSION   AJ001740
VERSION     AJ001740.1  GI:3152399
KEYWORDS    IgA protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 578)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Neisseria lactamica"
                     /isolate="NL3327"
                     /db_xref="taxon:486"
     CDS             <1. .>576
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA04965.1"
                     /db_xref="GI:3152400"
                     /db_xref="SPTREMBL:O69746"
                     /translation="GKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALV
                     GDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRL
                     HKFVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGAWLI
                     GGNTHMQGWGNNGVVSLSGDVRHANDYGPMPI"
     repeat_region   88. .95
                     /rpt_type=INVERTED
     repeat_region   139. .148
BASE COUNT      163 a    125 c    156 g    134 t
ORIGIN
```

Query Match 7.5%; Score 328.4; DB 1; Length 578;
Best Local Similarity 84.7%; Pred. No. 3.5e-66;
Matches 381; Conservative 0; Mismatches 66; Indels 3; Gaps 1;

```
Qy    184 ggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtc 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 60

Qy    244 ggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggcgtg 303
          |||||||| ||||| |||||||||||||||||||||||| ||||| ||||||||||||||
Db     61 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTG 120

Qy    304 gcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataacaat 363
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAAT 180

Qy    364 gttgattttggtgcggagggaagcaatcccgatcagcaccgctttttcttaccaaattgtg 423
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTTGATTTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG 240

Qy    424 aaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcatatg 483
          |||||||||| || ||| | |     ||    || |||||   |||||||| ||||||
Db    241 AAAAGAAATAACTACAAACCTG---ACAATTCACACCCTTACAACGGCGATTACCATATG 297

Qy    484 ccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatggat 543
          |||||  |||| || || || || ||||||||||||||| || |||||  || | |||
Db    298 CCGCGCCTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGG 357

Qy    544 gggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggcaga 603
          |||   |  ||  | |||   | | ||||| || || ||||| || |||||  ||||
Db    358 GGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACAC 417

Qy    604 caatattggcggtctgatgaagacgaaccc 633
          || |||||||| | |||||| ||| ||| |
Db    418 CACTATTGGCGTTATGATGATGACAAACAC 447
```

```
RESULT  16
NLIGA3354
LOCUS       NLIGA3354                584 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3354.
ACCESSION   AJ001739
VERSION     AJ001739.1  GI:3152397
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 584)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Neisseria lactamica"
                     /isolate="NL3354"
                     /db_xref="taxon:486"
     CDS             <1. .>582
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04964.1"
                     /db_xref="GI:3152398"
                     /db_xref="SPTREMBL:O69747"
                     /translation="FAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGD
                     QYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLHK
                     FVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDDKHGDLSYSGAWLIG
                     GNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGA"
     repeat_region   82. .88
                     /rpt_type=INVERTED
     repeat_region   134. .142
BASE COUNT      163 a    127 c    159 g    135 t
ORIGIN
```

Query Match 7.3%; Score 321.6; DB 1; Length 584;
Best Local Similarity 84.8%; Pred. No. 1.4e-64;
Matches 373; Conservative 0; Mismatches 64; Indels 3; Gaps 1;

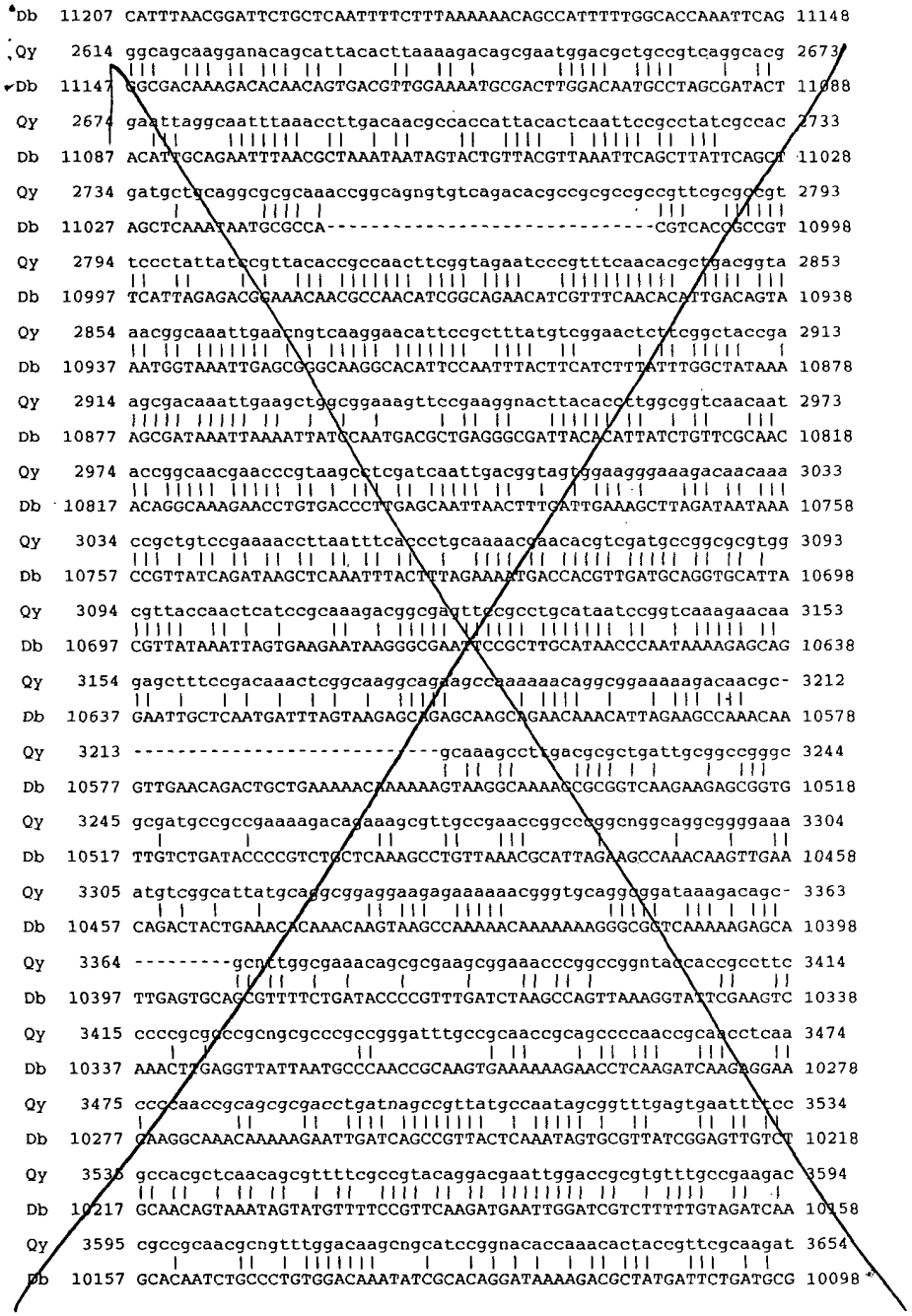
```
Qy    190 tttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaa 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAA 60

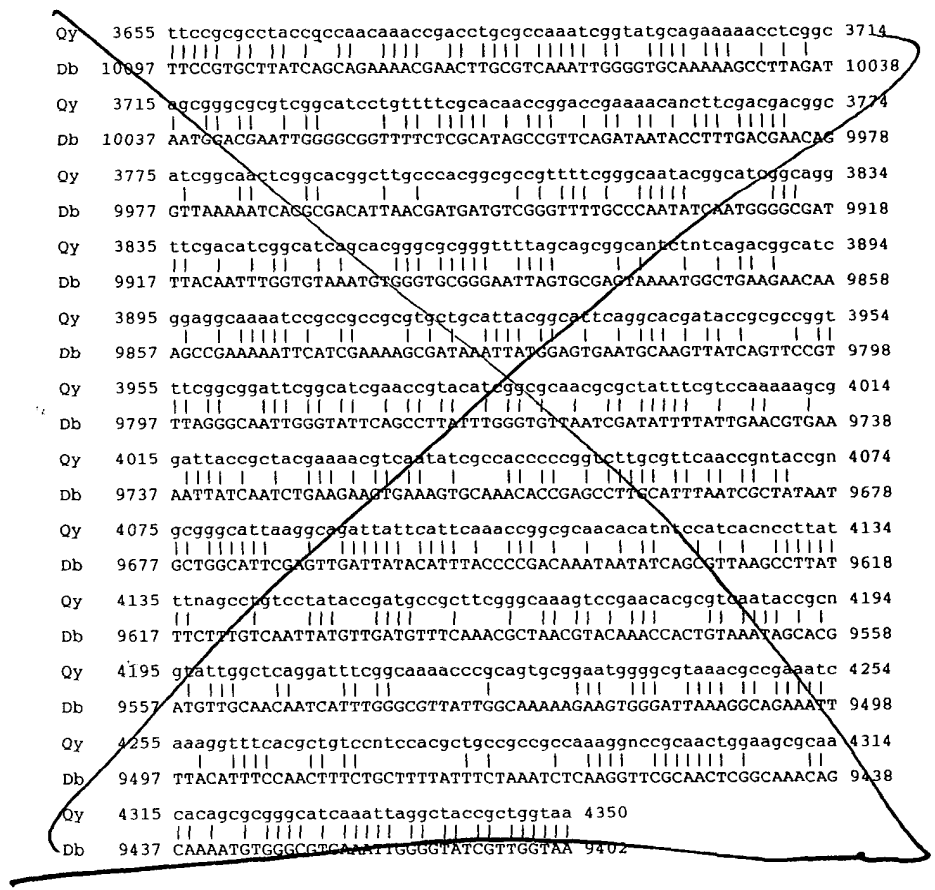
Qy    250 tcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggcgtggcggca 309
          || ||||| |||||||||||||||||||||||| ||||| |||||||||||||||||||
Db     61 TCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCA 120

Qy    310 ttggcgggcgatcaatatattgtgagcgtggcacataacggcggctataacaatgttgat 369
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGAT 180

Qy    370 tttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaattgtgaaaaga 429
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGA 240

Qy    430 aataattataaagcagggactaacggccatccttatggcggcgattatcatatgccgcgt 489
          ||||| || ||| | |     ||    || |||||   ||||||||| ||||||||||
```





RESULT 15
NLIGA3327
LOCUS NLIGA3327 578 bp DNA linear BCT 21-MAY-1998
DEFINITION Neisseria lactamica IgA protease homologue gene, isolate NL3327.
ACCESSION AJ001740
VERSION AJ001740.1 GI:3152399
KEYWORDS IgA protease.
SOURCE Neisseria lactamica.
ORGANISM Neisseria lactamica
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 578)
AUTHORS Sayers,J.R.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1997) Sayers J.R., Division of Molecular Medicine, University of Sheffield, Royal Hallamshire Hospital, Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE 2 (bases 1 to 578)
AUTHORS Sayers,J.R., Vitowski,S. and Read,R.
TITLE Strains of Neisseria lactamica carry an IgA1 protease gene homologue and some express IgA1 protease activity
JOURNAL Unpublished
FEATURES Location/Qualifiers

```
     source          1. .578
                     /organism="Neisseria lactamica"
                     /isolate="NL3327"
                     /db_xref="taxon:486"
     CDS             <1. .>576
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA04965.1"
                     /db_xref="GI:3152400"
                     /db_xref="SPTREMBL:O69746"
                     /translation="GKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALV
                     GDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRL
                     HKFVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGAWLI
                     GGNTHMQGWGNNGVVSLSGDVRHANDYGPMPI"
     repeat_region   88. .95
                     /rpt_type=INVERTED
     repeat_region   139. .148
BASE COUNT      163 a    125 c    156 g    134 t
ORIGIN
```

```
  Query Match             12.6%;  Score 550;  DB 1;  Length 578;
  Best Local Similarity   96.7%;  Pred. No. 4.5e-115;
  Matches  559;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy    184 ggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttggtc 243
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db      1 GGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 60

Qy    244 ggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtg 303
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     61 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTG 120

Qy    304 gcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaac 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAAT 180

Qy    364 gttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaattgtg 423
          |||||||||||||| || ||||| |||||||||||||||| ||||||||||||||||||||
Db    181 GTTGATTTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG 240

Qy    424 aaaagaaataattataagcctgacaattcacacccttacaacggcgattancatatgccg 483
          |||||||||||| || || |||||||||||||||||||||||||||||||||| |||||||||
Db    241 AAAAGAAATAACTACAAACCTGACAATTCACACCCTTACAACGGCGATTACCATATGCCG 300

Qy    484 cgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgaggggg 543
          ||  |||||||| || || || |||||||||||||||||||||||| ||||||||||||||||
Db    301 CGCCTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGG 360

Qy    544 aatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacaccac 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCAC 420

Qy    604 tattggcgttatgatgatgacaaacacggcgatttatcctactccggcgcatggttaatt 663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TATTGGCGTTATGATGATGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTAATT 480

Qy    664 ggcggcaatacacatatgcagggttggggaaataatggcgtanttagtttgagcggcgat 723
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    481 GGCGGCAATACACATATGCAGGGTTGGGGAAATAATGGCGTAGTTAGTTTGAGCGGCGAT 540

Qy    724 gtgcgccatgccaacgactatggccctatgccgattgc 761
          |||||||||||||||||||||||||||||||||||||
Db    541 GTGCGCCATGCCAACGACTATGGCCCTATGCCGATTGC 578
```

```
RESULT 10
NLIGA3354
LOCUS       NLIGA3354               584 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3354.
ACCESSION   AJ001739
VERSION     AJ001739.1  GI:3152397
```

```
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 584)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Neisseria lactamica"
                     /isolate="NL3354"
                     /db_xref="taxon:486"
     CDS             <1. .>582
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04964.1"
                     /db_xref="GI:3152398"
                     /db_xref="SPTREMBL:O69747"
                     /translation="FAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGD
                     QYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLHK
                     FVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDDKHGDLSYSGAWLIG
                     GNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGA"
     repeat_region   82. .88
                     /rpt_type=INVERTED
     repeat_region   134. .142
BASE COUNT      163 a    127 c    159 g    135 t
ORIGIN
```

```
  Query Match             12.4%;  Score 540;  DB 1;  Length 584;
  Best Local Similarity   96.2%;  Pred. No. 8.7e-113;
  Matches  562;  Conservative    0;  Mismatches   19;  Indels    3;  Gaps    1;

Qy    190 tttgcagtcgggggcgaaagatattgaggtntacaacaaaaaaggggagttggtcggcaaa 249
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      1 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAA 60

Qy    250 tcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggca 309
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     61 TCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCA 120

Qy    310 ttggtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgat 369
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    121 TTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGAT 180

Qy    370 tttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaattgtgaaaaga 429
          |||||||| || ||||| |||||||||||||||| ||||||||||||||||||||||||||
Db    181 TTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGA 240

Qy    430 aataattataagcctgacaattcacacccttacaacggcgattancatatgccgcgtttg 489
          |||||  || || |||||||||||||||||||||||||||||||||| ||||||||||  ||
Db    241 AATAACTACAAACCTGACAATTCACACCCTTACAACGGCGATTACCATATGCCGCGCCTG 300

Qy    490 cataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgaggggaatacc 549
          |||||  || || || |||||||||||||||||||||||||| ||||||||||||||||||
Db    301 CATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGAATACC 360

Qy    550 tattccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacaccactattgg 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTATTGG 420

Qy    610 cgttatgatgat---gacaaacacggcgatttatcctactccggcgcatggttaattggc 666
          ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2883 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 2942

Qy    632 gccccgagctggacagatcgggcaatgccgccgaagctttcaacggcactgcagatatcg 691
          || |||| ||| | | |||||||    |   ||   | ||||||||   || |    ||
Db   2943 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 3002

Qy    692 tcaaaaacatcatcggcgcggcaggagaaattgtcggc 729
             |||  | |    |   || || |||   || | |
Db   3003 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCGTTGAC 3040


RESULT  28
A96138
LOCUS       A96138                  1497 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 171 from Patent WO9924578.
ACCESSION   A96138
VERSION     A96138.1  GI:6779961
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1497)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 171 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1497
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      475 a    319 c    365 g    338 t
ORIGIN


  Query Match             7.3%;  Score 121.8;  DB 6;  Length 1497;
  Best Local Similarity   66.7%;  Pred. No. 1e-20;
  Matches  174;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
          | | | | | |  || ||||   |||||||||      ||  ||||| | |||||  ||
Db    243 CGAACGGACAGGATTTGAAGGTGTTATCGGCTATGAAACCCATTTTTCAGGACACGGACA 302

Qy    297 cgaagtccattccccttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
          ||||||  ||    || ||||| || ||||  ||| |   |  |||||||     ||| |
Db    303 CGAAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 362

Qy    357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
          |  | ||| || | | ||||||   | ||| ||   || |  |||   ||||||||
Db    363 CGTAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGC 422

Qy    417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
          |||||  |||||||||||  || ||||||||  |||||  |   ||  ||||  || ||||||||
Db    423 AGACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATAT 482

Qy    477 atacagctacgacataaaagg 497
          ||||||||||  | ||  |||||
Db    483 ATACAGCTACCATATCAAAGG 503


RESULT  29
A61829
LOCUS       [redacted]              1744 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 1 from Patent WO9711181.
ACCESSION   [redacted]
VERSION     A61829.1  GI:3715998
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1744)
  AUTHORS   Meyer,T.F., Rudel,T., Scheuerpflug,I., Fischer, Eckhard and
            Maier,J.
  TITLE     NUCLEIC ACID MOLECULES WHICH CODE PROTEINS WHICH MEDIATE THE
            ADHESION OF NEISSERIA CELLS TO HUMAN CELLS
  JOURNAL   Patent: WO 9711181-A 9 27-MAR-1997;
            MAX PLANCK GESELLSCHAFT (DE)
COMMENT     Other publication DE 19534579 970320.
FEATURES             Location/Qualifiers
     source          1. .1744
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
BASE COUNT      560 a    383 c    405 g    396 t
ORIGIN


  Query Match             7.3%;  Score 121.8;  DB 6;  Length 1744;
  Best Local Similarity   66.7%;  Pred. No. 1e-20;
  Matches  174;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
          | | | | | |  || ||||   |||||||||      ||  ||||| | |||||  ||
Db    284 CGAACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACA 343

Qy    297 cgaagtccattccccttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
          ||||||  ||    || ||||| || ||||  ||| |   |  |||||||     ||| |
Db    344 CGAAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 403

Qy    357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
          |  | ||| || | | ||||||   | ||| ||   || |  |||   ||||||||
Db    404 CGTAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGC 463

Qy    417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
          |||||  |||||||||||  || ||||||||  |||||  |   ||  ||||  || ||||||||
Db    464 AGACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATAT 523

Qy    477 atacagctacgacataaaagg 497
          ||||||||||  | ||  |||||
Db    524 ATACAGCTACCATATCAAAGG 544


RESULT  30
A61821
LOCUS       A61821                  3287 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 1 from Patent WO9711181.
ACCESSION   A61821
VERSION     A61821.1  GI:3715995
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 3287)
  AUTHORS   Meyer,T.F., Rudel,T., Scheuerpflug,I., Fischer, Eckhard and
            Maier,J.
  TITLE     NUCLEIC ACID MOLECULES WHICH CODE PROTEINS WHICH MEDIATE THE
            ADHESION OF NEISSERIA CELLS TO HUMAN CELLS
  JOURNAL   Patent: WO 9711181-A 1 27-MAR-1997;
            MAX PLANCK GESELLSCHAFT (DE)
COMMENT     Other publication DE 19534579 970320.
FEATURES             Location/Qualifiers
     source          1. .3287
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
                     /clone_lib="GENOMISCHE BIBLIOTHEK IN PBA"
                     /clone_lib="(B) CLON(E): H1967/PES25"
BASE COUNT     1017 a    740 c    752 g    778 t
ORIGIN


  Query Match             7.3%;  Score 121.8;  DB 6;  Length 3287;
  Best Local Similarity   66.7%;  Pred. No. 1.2e-20;
```

```
Db    425 ACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATATAT 484

Qy    479 acagctacgacataaaagg 497
          ||||||||  | ||  |||||
Db    485 ACAGCTACCATATCAAAGG 503


RESULT 30
A61829
LOCUS       A61829                  1744 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 9 from Patent WO9711181.
ACCESSION   A61829
VERSION     A61829.1  GI:3715998
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1744)
  AUTHORS   Meyer,T.F., Rudel,T., Scheuerpflug,I., Fischer, Eckhard and
            Maier,J.
  TITLE     NUCLEIC ACID MOLECULES WHICH CODE PROTEINS WHICH MEDIATE THE
            ADHESION OF NEISSERIA CELLS TO HUMAN CELLS
  JOURNAL   Patent: WO 9711181-A 9 27-MAR-1997;
            MAX PLANCK GESELLSCHAFT (DE)
COMMENT     Other publication DE 19534579 970320.
FEATURES             Location/Qualifiers
     source          1. .1744
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
BASE COUNT      560 a    383 c    405 g    396 t
ORIGIN


  Query Match             6.3%;  Score 119.8;  DB 6;  Length 1744;
  Best Local Similarity   66.4%;  Pred. No. 1.2e-17;
  Matches  172;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
          ||| | | |   |||||||  | ||||||  ||      ||  |||||  |  |||||  |||
Db    286 AACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACACG 345

Qy    299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
          ||  |  ||    ||  ||||| ||  ||||  |||  |   |  ||||||      |||  |
Db    346 AAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGCG 405

Qy    359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
            |  | |||  ||  |  |  ||||||    | |||  ||    ||  |  |||    |||||||| |
Db    406 TAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGCAG 465

Qy    419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
          ||||  ||||||||||| ||  |||||||| |||||  |  ||  ||||  ||  [redacted]
Db    466 ACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATATAT 525

Qy    479 acagctacgacataaaagg 497
          [redacted]  ||  |||||
Db    526 ACAGCTACCATATCAAAGG 544


RESULT 31
A61821
LOCUS       A61821                  3287 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 1 from Patent WO9711181.
ACCESSION   A61821
VERSION     A61821.1  GI:3715995
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 3287)
  AUTHORS   Meyer,T.F., Rudel,T., Scheuerpflug,I., Fischer, Eckhard and
            Maier,J.
  TITLE     NUCLEIC ACID MOLECULES WHICH CODE PROTEINS WHICH MEDIATE THE
            ADHESION OF NEISSERIA CELLS TO HUMAN CELLS
  JOURNAL   Patent: WO 9711181-A 1 27-MAR-1997;
            MAX PLANCK GESELLSCHAFT (DE)
COMMENT     Other publication DE 19534579 970320.
FEATURES             Location/Qualifiers
     source          1. .3287
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
                     /clone_lib="GENOMISCHE BIBLIOTHEK IN PBA"
                     /clone_lib="(B) CLON(E): H1967/PES25"
BASE COUNT     1017 a    740 c    752 g    778 t
ORIGIN


  Query Match             6.3%;  Score 119.8;  DB 6;  Length 3287;
  Best Local Similarity   66.4%;  Pred. No. 1.3e-17;
  Matches  172;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
          ||| | | |   |||||||  | ||||||  ||      ||  |||||  |  |||||  |||
Db   1829 AACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACACG 1888

Qy    299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
          ||  |  ||    ||  ||||| ||  ||||  |||  |   |  ||||||      |||  |
Db   1889 AAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGCG 1948

Qy    359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
            |  | |||  ||  |  |  ||||||    | |||  ||    ||  |  |||    |||||||| |
Db   1949 TAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGCAG 2008

Qy    419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
          ||||  ||||||||||| ||  |||||||| |||||  |  ||  ||||  ||  ||||||||||
Db   2009 ACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATATAT 2068

Qy    479 acagctacgacataaaagg 497
          ||||||||  | ||  |||||
Db   2069 ACAGCTACCATATCAAAGG 2087


RESULT 32
AF142582
LOCUS       AF142582                3294 bp    DNA     linear   BCT 18-MAY-1999
DEFINITION  Neisseria gonorrhoeae maf gene cluster, complete sequence.
ACCESSION   AF142582
VERSION     AF142582.1  GI:4838552
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 3294)
  AUTHORS   Eickenjaeger,S., Meyer,T.F., Fischer,E., Maier,J., Manning,P.A.,
            Rudel,T., Scheuerpflug,I., Schulz,E. and Schwan,E.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-1999) Molekulare Biologie, MPI fuer
            Infektionsbiologie, Monbijoustr. 2, Berlin 10117, Germany
FEATURES             Location/Qualifiers
     source          1. .3294
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
     gene            143. .457
                     /gene="mafI"
     CDS             143. .457
                     /gene="mafI"
                     /codon_start=1
                     /transl_table=11
                     /product="MafI"
                     /protein_id="AAD31037.1"
```

```
          Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .1356
                     /organism="Porphyromonas gingivalis"
                     /strain="W50"
                     /db_xref="taxon:837"
     CDS             1. .1356
                     /codon_start=1
                     /transl_table=11
                     /product="50 kD antigen PG1"
                     /protein_id="AAD33930.1"
                     /db_xref="GI:4929280"
                     /translation="MAKVIKTKKGLALNLKGKPLPEMLAEPAQSPTYAVVPDDFEGVI
                     PKVTARPGDKVRAGSALMHHKAYPEMKFTSPVSGEVIAVNRGAKRKVESIEVKPDGLN
                     EYESFPVGDPSALSAEQIKELLLSSGMWGFIKQRPYDIVATPDIAPRDIYITANFTAP
                     LAPDFDFIVRGEERALQTAIDALAKLTTGKVYVGLKPGSSLGLHNAEIVEVHGPHPAG
                     NVGVLINHTKPINRGETVWTLKATDLIVIGRFLLTGKADFTRMIAMTGSDAAAHGYVR
                     IMPGCNVFASFPGRLTIKESHERVIDGNVLTGKKLCEKEPFLSARCDQITVIPEGDDV
                     DELFGWAAPRLDQYSMSRAYFSWLQGKNKEYVLDARIKGGEEAMIMSNEYDRVFPMDI
                     YPEYLLKAIIAFDIDKMEDLGIYEVAPEDFATCEFVDTSKTELQRIVREGLDMLYKEM
                     N"
BASE COUNT      338 a    333 c    380 g    305 t
ORIGIN
```

Query Match 9.1%; Score 122.2; DB 1; Length 1356;
Best Local Similarity 46.0%; Pred. No. 4.2e-19;
Matches 622; Conservative 0; Mismatches 700; Indels 30; Gaps 6;

```
Qy       2 tgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcattt 61
Db      11 TTATAAAAACAAAAAAAGGCCTTGCACTTAATCTGAAAGGAAAACCGCTGCCCGAGATGC 70

Qy      62 atgacgggcccgtc---attaccgaagtcgcgttgcttggcgaagaatatgccggtatgc 118
Db      71 TGGCCGAACCGGCCCAAAGTCCTACTTACGCGGTCGTGCCCGACGATTTTGAAGGTGTTA 130

Qy     119 gcccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaag 178
Db     131 TCCCCAAGGTGACGGCTCGTCCGGGGGATAAGGTGCGTGCCGGCTCAGCACTGATGCACC 190

Qy     179 acaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatcc 238
Db     191 ACAAGGCATATCCGGAGATGAAGTTTACAAGTCCGGTTAGCGGCGAAGTGATCGCGGTGA 250

Qy     239 atcgcggcgaaaagcgcgtacttccagtcggtcgtgattgccgttgaaggcaacgacgaaa 298
Db     251 ATCGCGGTGCCAAGCGC-------AAGGTGTTGAGCATCGAGGTGAAACCGGACGGACTGA 304

Qy     299 tcgagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnn 358
Db     305 ACGAATACGAGTCATTCCCTGTCGGGGATCCGTCTGCCCTCTCTGCCGAACAGATCAAGG 364

Qy     359 gcaatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatcc 418
Db     365 AGCTTTTACTGTCGAGCGGTATGTGGGGTTTTATTAAGCAACGTCCTTACGACATAGTGG 424

Qy     419 ctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctng 478
Db     425 CTACACGGATATAGCTCCACGCGACATTTATATTACTGCCAACTTTACTGCACCATTGG 484

Qy     479 cggcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctgg 538
Db     485 CTCCGGACTTCGATTTCATCGTTCGAGGAGAAGAACGCGCCCTGCAGACTGCCATCGATG 544

Qy     539 tattgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgc 598
Db     545 CCTTGGCCAAACTCACGACAGGAAAGGTGTATGTGGGCCTGAAGCCGGGTTCATCTCTGG 604

Qy     599 cgtctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtt 658
Db     605 GCTTGCACAATGCAGAAATCGTAGAAGTACA---------CGGACCTCATCCGGCAGGTA 655
```

```
Qy     659 tgagtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggacca 718
Db     656 ACGTGGGCGTGCTGATCAATCATACGAAGCCAATCAATCGGGGCGAAACGGTGTGGACGC 715

Qy     719 tcaattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccg 778
Db     716 TCAAGGCTACCGACCTGATCGTGATCGGACGTTTCCTGCTTACGGGCAAAGCCGATTTTA 775

Qy     779 agcgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttt 838
Db     776 CCAGAATGATTGCCATGACCGGCTCAGACGCTGCAGCTCACGGATAGGTCCGTATTATGC 835

Qy     839 tgggtgcgaaagtatcgcaaattactgcgggcgaattg-----gttgacgcagacaacc 892
Db     836 CGGGTTGCAATGTCTTTGCTTCCTTCCCCGGCCGACTGACAATAAAGGAATCTCACGAGC 895

Qy     893 gcgtgatttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgg 952
Db     896 GTGTGATCGATGGCAATGTGCTGACCGGTAAGAAGCTCTGCGAGAAGGAGCCTTTCCTGT 955

Qy     953 gacgctaccacaatcagatttccgttatcgaagaaggccgc---agcaaagagctgttcg 1009
Db     956 CAGCCCGGTGTGACCAGATCACGGTGATCCCCGAAGGCGACGATGTGGACGAACTCTTCG 1015

Qy    1010 gctgggttgcgccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcc 1069
Db    1016 GGTGGGCTGCACCCCGTCTCGATCAGTACAGCATGAGCAGAGCTTATTTCTCTTGGTTGC 1075

Qy    1070 ---tgaaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatgg 1126
Db    1076 AGGGGAAAAACAAAGAGTACGTACTCGATGCCGGGATCAAGGGTGGCGAACGTGCTATGA 1135

Qy    1127 tgccgattggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgc 1186
Db    1136 TCATGAGCAACGAGTATGACCGCGTTTTCCCGATGGACATCTATCCGGAGTATTTGCTCA 1195

Qy    1187 gcgatttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacg 1246
Db    1196 AGGCTATTATAGCATTCGACATCGACAAGATGGAGGACTTAGGCATATATGAAGTGGCTC 1255

Qy    1247 aagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgt 1306
Db    1256 CGGAGGACTTTGCCACTTGCGAATTTGTGGATACATCCAAGATCGAGCTGCAGCGTATCG 1315

Qy    1307 tgcgtaaggtgctggaaaccnttgagaaggaa 1338
Db    1316 TTCGCGAGGGCTTGGATATGCTCTATAAGGAA 1347
```

```
RESULT 19
HIU20229
LOCUS       HIU20229                4926 bp    DNA     linear   BCT 09-FEB-1995
DEFINITION  Haemophilus influenzae BOLA (bolA), glutathione reductase (gor),
            phosphatidylserine decarboxylase (psd), 30K protein (rpmF), genes,
            complete cds.
ACCESSION   U20229
VERSION     U20229.1  GI:644850
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 4926)
  AUTHORS   Barcak,G.J. and Heimer,S.R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4926)
  AUTHORS   Barcak,G.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-1995) Gerard J. Barcak, Biochemistry, University
            of Maryland at Baltimore, 108 North Greene Street, Baltimore, MD
            21201, USA
FEATURES             Location/Qualifiers
     source          1. .4926
```

```
                /organism="Haemophilus influenzae"
                /strain="Rd"
                /isolate="KW20"
                /db_xref="taxon:727"
     CDS        complement(27. .224)
                /note="orf65"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62134.1"
                /db_xref="GI:644851"
                /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                VREGDVVKKGQVLFEDKKILV"
     gene       complement(519. .842)
                /gene="bolA"
     CDS        complement(519. .830)
                /gene="bolA"
                /function="stationary phase regulatory protein"
                /codon_start=1
                /transl_table=11
                /product="BOLA"
                /protein_id="AAA62135.1"
                /db_xref="GI:644852"
                /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                R"
     misc_feature   775. .785
                /note="encodes DNA uptake recognition sequence"
     RBS        complement(838. .842)
                /gene="bolA"
     protein_bind   877. .913
                /note="OxyR recongition sequence (Tartaglia, L.A., et al.,
                J. Mol Biol. 210: 709-710, 1989)"
                /bound_moiety="OxyR"
     RBS        915. .919
                /note="orf199"
     CDS        924. .1523
                /note="orf199"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62136.1"
                /db_xref="GI:644853"
                /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
     gene       1343. .3109
                /gene="gor"
     -35_signal 1343. .1348
                /gene="gor"
                /note="putative"
     -10_signal 1365. .1370
                /gene="gor"
                /note="putative"
     -35_signal 1585. .1590
                /gene="gor"
                /note="putative"
     -10_signal 1605. .1610
                /gene="gor"
                /note="putative"
     RBS        1728. .1731
                /gene="gor"
     CDS        1739. .3109
                /gene="gor"
                /EC_number="1.6.4.2"
                /function="reduces oxidized glutathione using NADPH"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="glutathione reductase"
                /protein_id="AAA62137.1"
                /db_xref="GI:644854"
                /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                EFVTMR"
     gene       complement(3218. .4090)
                /gene="psd"
     CDS        complement(3218. .4090)
                /gene="psd"
                /function="lipid metabolism"
                /codon_start=1
                /transl_table=11
                /product="phosphatidylserine decarboxylase"
                /protein_id="AAA62138.1"
                /db_xref="GI:644855"
                /translation="MMTSNSYWQRLKVAFQYVMPQIYLTQIAGWFAKQKWGKITHFVI
                KAFAKKYNIDMSIAQKEQFSDYASFNEFFIRPLKENARPINQNPTALCLPADGRISEC
                GHIDDNLLLQAKGHFFSLEDLLAEDKELVETFKNGEFVTTYLSPRDYHRVHMPCDATL
                RKMIYVPGDLFSVNPFLAQHVPNLFARNERVICVFDTEFGTMVQILVGATITASIGTT
                WAGVINPPRHNEVKTWTYEGESAVKLLKGQEMGWFQLGSTVINLFQANQVRLADHLSV
                NEPVRMGEILAYKK"
     misc_feature   complement(3364. .3374)
                /gene="psd"
                /note="encodes DNA uptake recognition sequence"
     misc_feature   complement(3780. .3790)
                /gene="psd"
                /note="encodes DNA uptake recognition sequence"
     misc_feature   3818. .3825
                /note="encodes DNA uptake recognition sequence"
     RBS        complement(4097. .4100)
                /note="psd"
     protein_bind   complement(4126. .4152)
                /note="OxyR recognition sequence (Toledano, M.B. et al.,
                Cell 78:897-909.)"
                /bound_moiety="OxyR"
     gene       4247. .4771
                /gene="rpmF"
     CDS        4247. .4771
                /gene="rpmF"
                /codon_start=1
                /transl_table=11
                /product="30K protein"
                /protein_id="AAA62139.1"
                /db_xref="GI:644856"
                /translation="MQKVKLPLTVDPIKDAQRRLDYVGYYAANQLERLAESVVNVLSD
                AQVTLSFYVDPQKLVVMKGKVQIDVDLECQRCNEPYKQTLECEFTYSPVANWDQADDL
                PEIYEPIEFNEFGEIDLIGTVEDELILALPLVPMHSSEHCEVSAQEQVFGVLPEELAK
                KPNPFAVLANLKQK"
     CDS        complement(4561. .>4926)
                /note="orf121"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62140.1"
                /db_xref="GI:644857"
                /translation="LQLRGGVSEFHRLLYLLIQQQLLKHRVNDAYHDVNEIFCFVEQP
                LAILLNLIYFCFKLANTANGFGFFANSSGNTPKTCSCADTSQCSDECIGTSGKAKISS
                SSTVPIKSISPNSLNSIGS"
BASE COUNT     1598 a    862 c    938 g   1528 t
ORIGIN


  Query Match             7.5%;  Score 101.4;  DB 1;  Length 4926;
  Best Local Similarity   69.6%;  Pred. No. 5e-14;
  Matches  149;  Conservative    0;  Mismatches   64;  Indels    1;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| ||| || |||||| |  | || || || ||||| | ||| |  ||||| ||
Db    224 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 165
```

```
Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           ||  |||   ||   | |   |||| ||| | || || || || |||| || ||||||
Db     164 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 105

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           || |  ||||| ||    |||||||||| || |||||||| ||||| || ||||||||||
Db     104 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTTGAAGAC 45

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccng 214
           |||||  |||| || ||  | ||||| || || |
Db      44 AAAAA-AATCCTGGTGTAATTTTTACAGCCCCTG 12


RESULT  20
AF429315/c
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581 .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN


  Query Match             3.4%;  Score 45.6;  DB 9;  Length 125020;
  Best Local Similarity   10.7%;  Pred. No. 2.1;

  Matches   97;  Conservative  374;  Mismatches  431;  Indels    7;  Gaps    2;

Qy     339 aagcggcgangaantnngnngcaatctgatccaatccggtttgtggactgcgctgcgtan 398
           | : :: | |   :  |  : :::: :    :|   :  : ::|::  :
Db   17784 AGSKSSRGGTGYYMKKKGGGKGSMSSKKWKGSSTSRRRGSSAKSCSYMWGMSMCMSCMSM 17725

Qy     399 ccgtccgttcagcaaaatccctgccgtcgatgccgagccgttcgccatcttcgtcaatgc 458
           :::: : ::  :    :::::|:: :|: : :   :| :::::   |::| :: :
Db   17724 AKSYMMCYCYMYRMSSYMSYYYKCYSCMGMSSTSYSYSCCWKMSSWGSCYKCMKYYSGWS 17665

Qy     459 gatggacaccaatccgctngcggcagaccctgtggttgtgatcaaagaagccgngncganga 518
           :   ::  ::   | ::: : ::: :   :::::::: ::  :    :| :::   |: ::
Db   17664 SYSTSMGYYSSTSCKKYKCSWSSMYKCKTSKSYRRKRSYYYWGGGKRAKKKYYCAGRRRR 17605

Qy     519 tttcagacgangtntgctggtattgagccgtttgaccgagcgtaaaatccatgtgtgtaa 578
             :: |      : |::  :|:  : : |:::  |  : |: : |    |::  :
Db   17604 -MSYWKCCAKWWMSYCCWSYCMTYYYSKSCTYKSSTCYKRGGYYWCSKTCYSAGGKSRSM 17546

Qy     579 ggcagctggcgcagacgtgccgtctgaaaatgctgccaacatccgaaacacatgaattcgg 638
             |:   :::::::::  :::  :   |   :::::  : | :         ||:   |::
Db   17545 YYCMMRSSKSSSWSMSMARSSWCMGWGAGYRRSKRSAGWGAGNRSSKGKRSTGMKRACSK 17486

Qy     639 cggcccgcatccggccggtttgagtggcacgcacattcatttcattgagccggtcggtgc 698
             |:: :   :::  |  ::::| :|    | |:::: ::       : :: :  | :::
Db   17485 KTGSYGSTGRSMKKKKGYSKYSRGMKGKKKTCYCMWKYYKYRKTSMCWWYYMKSWGYKRY 17426

Qy     699 aaacaaaaccgtttggaccatcaattatcaagatgtaattgccatcggacgtttgtttgc 758
           : |:    |  | :   :  : :: : ::|  :::|  ::   | :: :::::
Db   17425 KRCCMKKKGCTGYRGMSSKKSGYKSMRGMRGSSYSTSCWKSCWGYSWMKCMKYMSYKKKR 17366

Qy     759 aacaggccgtctgaacaccgagcgcgtgattgctttgggtggttctcaagtcaacaaacc 818
           : ::|::::: ::         :::   :: :: :: :  ::    : |:  : :   :
Db   17365 RSMRGSMSSKGWRGYAGRGCYSSSMWSTRKRSKCYSYKSYKKGRGKMKGWGGMKRGSKY 17306

Qy     819 acgcctcttgcgtaccgttttgggtgcgaaagtatcgcaaattactgcgggcgaattggt 878
           :::: ::: ::   |::|: :::| : :  :| | : |    :: |::  :|  :::
Db   17305 WSSMKKMKRSSWSKCYSTKSYKSGRRKSKGWGRSTKSKAKSSMRMAGSKCTYGSSYWSNN 17246

Qy     879 tgacgcagacaaccgcgtgatttccggttcggtattgaacggcgcgatta------caca 932
           :  :: :    ::  : :|          |   |  | | ||  |       :|::
Db   17245 RRNNRMGKTGCNYMYRRSRAMMNGNAAAAGCTTCCCCANTNGGGGGAAAAAGGGCGSASR 17186

Qy     933 aggcgcgcacgattatttgggacgctaccacaatcagatttccgttatcgaagaaggccg 992
           |: :: :: ::  ::  |::| :::    :: :::|:|  :|  :  ::| |  :|||
Db   17185 ASCYKGRMSSKSCYRSGTRRRCMKSKCRRGSRGKSMGMTRGRSGGKTSYSAKGSGRGCYY 17126

Qy     993 cagcaaagagctgttcggctgggttgcgccgcagccggacaaatactccatcacgcgtac 1052
           |:|  : |::| :    :::::: : : : : |:::| :::|::   :|  |:|::   |
Db   17125 CWGWGRKGRKCMSSRKMKMYKSYYRRKRWMTCMKMCYSMYMAMYCRSMCCMCMCKSCCGC 17066

Qy    1053 gaccctcggccatttcctgaaaaacaaactcttcaagttcacgacagccgtcaacggtgg 1112
           : : : ::: :    :  ::::    :|  |::  :    | ::   :: |:||
Db   17065 YSMGMSSYSYSGKYSWGMKSYMWRSYYSKRSTSKAWRSSKRGMGTGGRYKGGGRSYGKGG 17006

Qy    1113 cgaccgcgccatggtgccgattggtacttacgagcgcgtaatgccgctagacatcctgcc 1172
           :   | :   :||::  |   |:: ::: ::   ::: : :: | :: :|: :::
Db   17005 GSYKGGGSWGGKGGKSTRGSSAGKKSSYKKMSCCARYKSMSYSKCYSRWMWCMSYYYCWG 16946

Qy    1173 taccctgcttttgcgcgatttaatcgtcggcgataccgacagcgcgcaagcattgggttg 1232
             |:   :   : :::   :|  ::: :|:: |: :: : : ::: : :  ::: :
Db   16945 GKQWSSCWWSSMRKKSKGGSWAKGMVWDKGSVSTDKSDKMBSRSBSKVKSKWSMSRRGKY 16886

Qy    1233 cttggaatt 1241
           ::  : :::
Db   16885 YYCTDSMKK 16877


RESULT  21
BACGNTZA/c
LOCUS       BACGNTZA               36200 bp    DNA     linear   BCT 05-FEB-1999
DEFINITION  Bacillus subtilis 36kb sequence between gntZ and trnY genes
            encoding 34 ORFs.
ACCESSION   D78193
```

```
          Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .1356
                     /organism="Porphyromonas gingivalis"
                     /strain="W50"
                     /db_xref="taxon:837"
     CDS             1. .1356
                     /codon_start=1
                     /transl_table=11
                     /product="50 kD antigen PG1"
                     /protein_id="AAD33930.1"
                     /db_xref="GI:4929280"
                     /translation="MAKVIKTKKGLALNLKGKPLPEMLAEPAQSPTYAVVPDDFEGVI
                     PKVTARPGDKVRAGSALMHHKAYPEMKFTSPVSGEVIAVNRGAKRKVLSIEVKPDGLN
                     EYESFPVGDPSALSAEQIKELLLSSGMWGFIKQRPYDIVATPDIAPRDIYITANFTAP
                     DAPDFDFIVRGEERALQTAIDALAKLTTGKVYVGLKPGSSLGLHNAEIVEVHGPHPAG
                     NVGVLINHTKPINRGETVWTLKATDLIVIGRFLLTGKADFTRMLAMTGSDAAAHGYVR
                     IMPSCNVFASPPGRLTIKESHERVIDGNVLTGKKLCEKEPFLSARCDQITVIPEGDDV
                     DELFGWAAPRLDQYSMSRAYFSWLQGKNKEYVLDARIKGGERAMIMSNEYDRVFPMDI
                     YPEYLLDKAIIAFDIDKMEDLGIYEVAPEDFATCEFVDTSKIELQRIVREGLDMLYKEM
                     N"
BASE COUNT      338 a    333 c    380 g    305 t
ORIGIN
```

Query Match 8.5%; Score 113.6; DB 1; Length 1356;
Best Local Similarity 46.1%; Pred. No. 3.6e-17;
Matches 623; Conservative 0; Mismatches 699; Indels 30; Gaps 6;

```
Qy       2 tgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgttt 61
Db      11 TTATAAAAACAAAAAAAGGCCTTGCACTTAATCTGAAAGGAAAACCGCTGCCCGAGATGC 70

Qy      62 acgacggcccggcc---attaccgaagtcgcgttgcttggcgaagaatatgccggtatgc 118
Db      71 TGGCCGAACCGGCCCAAAGTCCTACTTACGCGGTCGTGCCCGACGATTTTGAAGGTGTTA 130

Qy     119 gcccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaag 178
Db     131 TCCCCAAGGTGACGGCTCGTCCGGGGGATAAGGTGCGTGCCGGCTCAGCACTGATGCACC 190

Qy     179 acaaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccggattc 238
Db     191 ACAAGGCATATCCGGAGATGAAGTTTACAAGTCCGGTTAGCGGCGAAGTGATCGCGGTGA 250

Qy     239 accgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaa 298
Db     251 ATCGCGGTGCCAAGCGCAAGGTGTTGAGCATCGAGGTGAAACCGGACGGACTGAACGAAT 310

Qy     299 tcgagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgcc 358
Db     311 ACGAGT------CATTCCCTGTCGGGGATCCGTCTGCCCTCTCTGCCGAACAGATCAAGG 364

Qy     359 gcaacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattc 418
Db     365 AGCTTTTACTGTCGAGCGGTATGTGGGGTTTTATTAAGCAACGTCCTTACGACATAGTGG 424

Qy     419 ctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctgg 478
Db     425 CTACACCGGATATAGCTCCACGCGACATTTATATTACTGCCAACTTTACTGCACCATTGG 484

Qy     479 ctgccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttgg 538
Db     485 CTCCGGACTTCGATTTCATCGTTCGAGGAGAAGAACGCGCCCTGCAGACTGCCATCGATG 544

Qy     539 tattgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgc 598
Db     545 CGTTGGCCAAACTCACGACAGGAAAGGTGTATGTGGGCCTGAAGCCGGGTTCATCTCTGG 604

Qy     599 cgtctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtt 658
Db     605 GCTTGCACAATGCAGAAATCGTAGAAGTACA---------CGGACCTCATCCGGCAGGTA 655

Qy     659 tgagtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggacca 718
Db     656 ACGTGGGCGTGCTGATCAATCATACGAAGCCAATCAATCGGGGCGAAACGGTGTGGACGC 715

Qy     719 tcaattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccg 778
Db     716 TCAAGGCTACCGACCTGATCGTGATCGGACGTTTCCTGCTTACGGGCAAAGCCGATTTTA 775

Qy     779 agcgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttt 838
Db     776 CCAGAATGATTGCCATGACCGGCTCAGACGCTGCAGCTCACGGATACGTCCGTATTATGC 835

Qy     839 tgggtgcgaaagtatcgcaaattactgcgggcgaattg------gttgacacagacaacc 892
Db     836 CGGGTTGCAATGTCTTTGCTTCCTTCCCCGGCCGACTGACAATAAAGGAATCTCACGAGC 895

Qy     893 gcgtgatttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgg 952
Db     896 GTGTGATCGATGGCAATGTGCTGACCGGTAAGAAGCTCTGCGAGAAGGAGCCTTTCCTGT 955

Qy     953 gacgctaccacaatcagatttccgttatcgaagaaggccgc---agcaaagagctgttcg 1009
Db     956 CAGCCCGGTGTGACCAGATCACGTTGATCCCCGAAGGCGACGATGTGGACGAACTCTTCG 1015

Qy    1010 gctgggttgcgccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcc 1069
Db    1016 GGTGGGCTGCACCCCGTCTCGATCAGTACAGCATGAGCAGAGCTTATTTCTCTTGGTTGC 1075

Qy    1070 ---tgaaaaacaaactcttcaagttcaacacagccgtcaacggcggcgaccgcgccatgg 1126
Db    1076 AGGGGAAAAACAAAGAGTACGTACTCGATGCCCGGATCAAGGGTGGCGAACGTGCTATGA 1135

Qy    1127 tgccgattggtacttacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgc 1186
Db    1136 TCATGAGCAACGAGTATGACCGCGTTTTCCCGATGGACATCTATCGGAGTATTTGCTCA 1195

Qy    1187 gcgatttaatcgtcggcgataccgacagcgcgcaggcattgggttgcttggaattggacg 1246
Db    1196 AGGCTATTATAGCATTCGACATCGACAAGATGGAGGACTTAGGCATATATGAAGTGGCTC 1255

Qy    1247 aagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgt 1306
Db    1256 CGGAGGACTTTGCCACTTGCGAATTTGTGGATACATCCAAGATCGAGCTGCAGCGTATCG 1315

Qy    1307 tgcgcaaagtgctggaaaccattgagaaggaa 1338
Db    1316 TGCGCGAGGGCTTGGATATGCTGCTATAAGGAA 1347
```

```
LOCUS       HIU20229                4926 bp    DNA     linear   BCT 09-FEB-1995
DEFINITION  Haemophilus influenzae BOLA (bolA), glutathione reductase (gor),
            phosphatidylserine decarboxylase (psd), 30K protein (rpmF), genes,
            complete cds.
ACCESSION   U20229
VERSION     U20229.1  GI:644850
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 4926)
  AUTHORS   Barcak,G.J. and Heimer,S.R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4926)
  AUTHORS   Barcak,G.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-1995) Gerard J. Barcak, Biochemistry, University
            of Maryland at Baltimore, 108 North Greene Street, Baltimore, MD
            21201, USA
FEATURES             Location/Qualifiers
     source          1. .4926
```

```
                /organism="Haemophilus influenzae"
                /strain="Rd"
                /isolate="KW20"
                /db_xref="taxon:727"
CDS             complement(27. .224)
                /note="orf65"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62134.1"
                /db_xref="GI:644851"
                /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                VREGDVVKKGQVLFEDKKILV"
gene            complement(519. .842)
                /gene="bolA"
CDS             complement(519. .830)
                /gene="bolA"
                /function="stationary phase regulatory protein"
                /codon_start=1
                /transl_table=11
                /product="BOLA"
                /protein_id="AAA62135.1"
                /db_xref="GI:644852"
                /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                R"
misc_feature    775. .785
                /note="encodes DNA uptake recognition sequence"
RBS             complement(838. .842)
                /gene="bolA"
protein_bind    877. .913
                /note="OxyR recongition sequence (Tartaglia, L.A., et al.,
                J. Mol Biol. 210: 709-710, 1989)"
                /bound_moiety="OxyR"
RBS             915. .919
                /note="orf199"
CDS             924. .1523
                /note="orf199"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62136.1"
                /db_xref="GI:644853"
                /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
gene            1343. .3109
                /gene="gor"
-35_signal      1343. .1348
                /gene="gor"
                /note="putative"
-10_signal      1365. .1370
                /gene="gor"
                /note="putative"
-35_signal      1585. .1590
                /gene="gor"
                /note="putative"
-10_signal      1605. .1610
                /gene="gor"
                /note="putative"
RBS             1728. .1731
                /gene="gor"
CDS             1739. .3109
                /gene="gor"
                /EC_number="1.6.4.2"
                /function="reduces oxidized glutathione using NADPH"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="glutathione reductase"
                /protein_id="AAA62137.1"
                /db_xref="GI:644854"
                /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                EFVTMR"
gene            complement(3218. .4090)
                /gene="psd"
CDS             complement(3218. .4090)
                /gene="psd"
                /function="lipid metabolism"
                /codon_start=1
                /transl_table=11
                /product="phosphatidylserine decarboxylase"
                /protein_id="AAA62138.1"
                /db_xref="GI:644855"
                /translation="MMTSNSYWQRLKVAFQYVMPQIYLTQIAGWFAKQKWGKITHFVI
                KAFAKKYNIDMSIAQKEQFSDYASFNEFFIRPLKENARPINQNPTALCLPADGRISEC
                GHIDDNLLLQAKGHFFSLEDLLAEDKELVETFKNGEFVTTYLSPRDYHRVHMPCDATL
                RKMIYVPGDLFSVNPFLAQHVPNLFARNERVICVFDTEFGTMVQILVGATITASIGTT
                WAGVINPPRHNEVKTWTYEGESAVKLLKGQEMGWFQLGSTVINLFQANQVRLADHLSV
                NEPVRMGEILAYKK"
misc_feature    complement(3364. .3374)
                /gene="psd"
                /note="encodes DNA uptake recognition sequence"
misc_feature    complement(3780. .3790)
                /gene="psd"
                /note="encodes DNA uptake recognition sequence"
misc_feature    3818. .3825
                /note="encodes DNA uptake recognition sequence"
RBS             complement(4097. .4100)
                /note="psd"
protein_bind    complement(4126. .4152)
                /note="OxyR recognition sequence (Toledano, M.B. et al.,
                Cell 78:897-909.)"
                /bound_moiety="OxyR"
gene            4247. .4771
                /gene="rpmF"
CDS             4247. .4771
                /gene="rpmF"
                /codon_start=1
                /transl_table=11
                /product="30K protein"
                /protein_id="AAA62139.1"
                /db_xref="GI:644856"
                /translation="MQKVKLPLTVDPIKDAQRRLDYVGYYAANQLERLAESVVNVLSD
                AQVTLSFYVDPQKLVVMKGKVQIDVDLECQRCNEPYKQTLECEFTYSPVANWDQADDL
                PEIYEPIEFNEFGEIDLIGTVEDELILALPLVPMHSSEHCEVSAQEQVFGVLPEELAK
                KPNPFAVLANLKQK"
CDS             complement(4561. .>4926)
                /note="orf121"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62140.1"
                /db_xref="GI:644857"
                /translation="LQLRGGVSEFHRLLYLLIQQQLLKHRVNDAYHDVNEIFCFVEQP
                LAILLNLIYFCFKLANTANGFGFFANSSGNTPKTCSCADTSQCSDECIGTSGKAKISS
                SSTVPIKSISPNSLNSIGS"
BASE COUNT      1598 a    862 c    938 g   1528 t
ORIGIN


  Query Match             7.5%;  Score 100.6;  DB 1;  Length 4926;
  Best Local Similarity   68.6%;  Pred. No. 5.8e-14;
  Matches  153;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
           ||||||| ||| || |||||| |  | || || || ||||| | ||| |  ||||   |
Db     224 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 165
```

```
Qy      61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           |   ||||   ||  | |   |||| ||| | || || || || ||||  || ||||
Db     164 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 105

Qy     121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           || || ||||| ||    |||||||||| || |||||||| ||||| || [redacted]
Db     104 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTGAAGAC 45

Qy     181 aaaacgaatccgggcgtggtgtttactgcgccggcttcaggca 223
           [redacted] ||||| || ||  | ||||| || || ||    || |
Db      44 AAAAA-AATCCTGGTGTAATTTTTACAGCCCCTGCAAGCGGTA 3
```

```
RESULT 20
AF429315
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN
```

```
Query Match             3.5%;  Score 47.4;  DB 9;  Length 125020;
Best Local Similarity   11.4%;  Pred. No. 0.76;
```

```
  Matches   45;  Conservative  172;  Mismatches  177;  Indels    0;  Gaps    0;

Qy     239 accgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaa 298
           ::: ::::: : : |: :  ::   : | | :|:|  :::| :  :::   :: :: :
Db   17040 RSYWKRSMKCWSRMCSRSRSSKCKSRGCGGSMGKGKGGKSYGRKTKRKSRGKMKGAKWYM 17099

Qy     299 tcgagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgcc 358
           :::::  :  ::::: : |::|: : : ||: :|:     :  :  :   : : ::| :
Db   17100 YYRRSMRKMKMYSSKGMYCMYCWCWGRRGCYCSCMTSRSAMCCSYCYAKCKSMCYSCYYG 17159

Qy     359 gcaacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattc 418
           ::  :: | :::  ::::: : :  :|:|||       ||  |   |  | ||  ||
Db   17160 MSMKGYYYACSYRGSMSSKYCMRGSTYSTSCGCCCTTTTTCCCCCNANTGGGGAAGCTTT 17219

Qy     419 ctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctgg 478
           ::|:: ::      : ::  ::  :: ::  |  :|  ::: : :::| :| | :
Db   17220 TNCNKKTYSYYRKRNGCAMCKYNNYYNNSWRSSCRAGMSCTKYKSSMTMSMASYCWCMSM 17279

Qy     479 ctgccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttgg 538
           :: ::: : :  ::  :  ::    ::: :::|| :: : ::: ::::|:::  : :
Db   17280 YYCSMRSMASRGMSWSSYMKMMKSSWRMSCYMKCCWCMKMCYCMMRSMRSRGMSYYMYAS 17339

Qy     539 tattgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgc 598
           :  ::||: |  :|::  : :     ::: :  :|  :::|:|:  |: ::: | : :
Db   17340 WKSSSRGCYCTRCYWCMSSKSCYKSYYMMMRSKRMKGMKWSRCWGSMWGSASRSSCYKCY 17399

Qy     599 cgtctgaaaatgctgccaacatcgaaacacatga 632
           :  :::  : :    :::: | :::: :::
Db   17400 KSMRCSMMSSKCYRCAGCMMMKGGYMRYMRCWSM 17433
```

```
RESULT 21
BACGNTZA/c
LOCUS       BACGNTZA                36200 bp    DNA     linear   BCT 05-FEB-1999
DEFINITION  Bacillus subtilis 36kb sequence between gntZ and trnY genes
            encoding 34 ORFs.
ACCESSION   D78193
VERSION     D78193.1  GI:1064780
KEYWORDS    phospho-bete-glucosidase; 6-phosphogluconate dehydrogenase; alkyl
            hydroperoxide reductase; amino acid permease; arginase; orthinine
            aminotransferase; positive regulatory protein; gntZ; ahp; bglA;
            yyd; yyc; formaldehyde dehydrogenase.
SOURCE      Bacillus subtilis (strain:168) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 36200)
  AUTHORS   Ogasawara,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1995) Naotake Ogasawara, Nara Institute of
            Science and Technology, Graduate School of Biological Sciences;
            8916-5 Takayama-cho, Ikoma, Nara 630-01, Japan
            (E-mail:nogasawa@bs.aist-nara.ac.jp, Tel:07437-2-5430,
            Fax:07437-2-5439)
REFERENCE   2  (sites)
  AUTHORS   Zhang,J. and Aronson,A.
  TITLE     A Bacillus subtilis bglA gene encoding phospho-beta-glucosidase is
            inducible and closely linked to a NADH dehydrogenase-encoding gene
  JOURNAL   Gene 140 (1), 85-90 (1994)
  MEDLINE   94171085
REFERENCE   3  (sites)
  AUTHORS   Hartford,O.M. and Dowds,B.C.
  TITLE     Isolation and characterization of a hydrogen peroxide resistant
            mutant of Bacillus subtilis
  JOURNAL   Microbiology 140 (Pt 2), 297-304 (1994)
  MEDLINE   94236234
REFERENCE   4  (sites)
  AUTHORS   Ogasawara,N., Nakai,S. and Yoshikawa,H.
  TITLE     Systematic sequencing of the 180 kilobase region of the Bacillus
            subtilis chromosome containing the replication origin
  JOURNAL   DNA Res. 1 (1), 1-14 (1994)
  MEDLINE   96051385
REFERENCE   5  (sites)
```

```
          Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .1356
                     /organism="Porphyromonas gingivalis"
                     /strain="W50"
                     /db_xref="taxon:837"
     CDS             1. .1356
                     /codon_start=1
                     /transl_table=11
                     /product="50 kD antigen PG1"
                     /protein_id="AAD33930.1"
                     /db_xref="GI:4929280"
                     /translation="MAKVIKTKKGLALNLKGKPLPEMLAEPAQSPTYAVVPDDFEGVI
                     PKVTARPGDKVRAGSALMHHKAYPEMKFTSPVSGEVIAVNRGAKRKVLSIEVKPDGLN
                     EYESFPVGDPSALSAEQIKELLLSSGMWGFIKQRPYDIVATPDIAPRDIYITANFTAP
                     LAPDFDFIVRGEERALQTAIDALAKLTTGKVYVGLKPGSSLGLHNAEIVEVHGPHPAG
                     NVGVLINHTKPINRGETVWTLKATDLIVIGRFLLTGKADFTRMIAMTGSDAAAHGYVR
                     IMPGCNVFASFPGRLTIKESHERVIDGNVLTGKKLCEKEPFLSARCDQIFVIPEGDDV
                     DELFGWAAPRLDQYSMSRAYFSWLQGKNKEYVLDARIKGGERAMIMSNEYDRVFPMDI
                     YPEVLLKAIIAFDIDKMEDLGIYEVAPEDFATCEFVDTSKIELQRIVREGLDMLYKEM
                     N"
BASE COUNT      338 a    333 c    380 g    305 t
ORIGIN
```

```
  Query Match           9.4%;  Score 126.4;  DB 1;  Length 1356;
  Best Local Similarity  46.7%;  Pred. No. 4.4e-21;
  Matches  631;  Conservative    0;  Mismatches  691;  Indels   30;  Gaps    6;

Qy       2 tgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagacggagcaagtcattt 61
          | || ||||  |||||||| ||   ||      |  || | ||||  ||  |  ||
Db      11 TTATAAAAACAAAAAAAGGCCTTGCACTTAATCTGAAAGGAAAACCGCTGCCCGAGATGC 70

Qy      62 atgacggcccggcc---attaccgaagtcgcgttgcttggcgaagaatatgtcggcatgc 118
             | ||  ||||||   | | |     |||| |  |  ||| || || ||  ||  |
Db      71 TGGCCGAACCGGCCCAAAGTCCTACTTACGCGGTCGTGCCCGACGATTTTGAAGGTGTTA 130

Qy     119 gcccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaag 178
           ||||  | |||          || ||    ||      ||| ||  ||| |  |
Db     131 TCCCCAAGGTGACGGCTCGTCCGGGGGATAAGGTGCGTGCCGGCTCAGCACTGATGCACC 190

Qy     179 acaaaaagaatccgggcgtagtatttactgtgccggcttcaggcaaaatcgccgctattc 238
           ||||     ||||||    |    ||||||    |||| |  ||| || |   |||  |
Db     191 ACAAGGCATATCCGGAGATGAAGTTTACAAGTCCGGTTAGCGGCGAAGTGATCGCGGTGA 250

Qy     239 accgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaa 298
           | || || |  ||||||    |  |    ||| | |      || ||     |||||
Db     251 ATCGCGGTGCCAAGCGCAAGGTGTTGAGCATCGAGGTGAAACCGGACGGACTGAACGAAT 310

Qy     299 tcgagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgcc 358
           |||||    |   ||  |     | | |      |    ||||  |  |
Db     311 ACGAGTCATTCCCTGTCGGGGAT------CCGTCTGCCCTCTCTGCCGAACAGATCAAGG 364

Qy     359 gcaacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatcc 418
             |  |      ||  | |||  |    ||      |||||  | |  ||  | |
Db     365 AGCTTTTACTGTCGAGCGGTATGTGGGGTTTTATTAAGCAACGTCCTTACGACATAGTGG 424

Qy     419 ctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctgg 478
           || |    |||  |  ||  || ||| |  | | ||| |     ||    ||  |||
Db     425 CTACACCGGATATAGCTCCACGCGACATTTATATTACTGCCAACTTTACTGCACCATTGG 484

Qy     479 ctgccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttgg 538
           || | |||       |||||  |   || ||  |      | | |   |    | | |
Db     485 CTCCGGACTTCGATTTCATCGTTCGAGGAGAAGAACGCGCCCTGCAGACTGCCATCGATG 544

Qy     539 tattgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgc 598
           |||  |   || ||   | | ||  | ||||| |          | ||  ||    ||
Db     545 CCTTGGCCAAACTCACGACAGGAAAGGTGTATGTGGGCCTGAAGCCGGGTTCATCTCTGG 604

Qy     599 cgtctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggct 658
             |   | |||| |  |   | |||  |||          ||| || ||||| || ||
Db     605 GCTTGCACAATGCAGAAATCGTAGAAGTACA---------CGGACCTCATCCGGCAGGTA 655
```

```
Qy     659 tgagtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggacca 718
           |||  ||  ||  ||   |  ||||| ||      |    |||| ||||||||
Db     656 ACGTGGGCGTGCTGATCAATCATACGAAGCCAATCAATCGGGGCGAAACGGTGTGGACGC 715

Qy     719 tcaattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccg 778
           ||||   |  ||| |||| |  ||||||||||  |  | || |||        ||
Db     716 TCAAGGCTACCGACCTGATCGTGATCGGACGTTTCCTGCTTACGGGCAAAGCCGATTTTA 775

Qy     779 agcgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgggtaccgttt 838
           |  || |||||  ||  ||||   | |        |   | | ||||   |
Db     776 CCAGAATGATTGCCATGACCGGCTCAGACGCTGCAGCTCACGGATACGTCCGTATTATGC 835

Qy     839 tgggtgcgaaggtgtctcaacttaccgccggcgaattggtt------gacgcggacaacc 892
           ||||   || || | |        | ||||| | ||          ||  | || | |
Db     836 CGGGTTGCAATGTCTTTGCTTCCTTCCCCGGCCGACTGACAATAAAGGAATCTCACGAGC 895

Qy     893 gcgtgatttccggttcggtattgaacggtgcgattgcagaaggcgcgcatgattatttgg 952
           | ||||||    ||    ||  ||| |||  ||        |  |  ||   ||
Db     896 GTGTGATCGATGGCAATGTGCTGACCGGTAAGAAGCTCTGCGAGAAGGAGCCTTTCCTGT 955

Qy     953 gacgctaccacaatcagatttccgttatcgaagaaggccgcagca---aagagctgttcg 1009
           |  |        |||||  | || |||  ||||||  |       | || || ||||
Db     956 CAGCCCGGTGTGACCAGATCACGGTGATCCCCGAAGGCGACGATGTGGACGAACTCTTCG 1015

Qy    1010 gctgggttgcgccgcagccggacaaatactccatcacgcgcaccactctcggccatttcc 1069
           | |||| ||| || |  |  || ||| |||  |  |  |  | ||      || |
Db    1016 GGTGGGCTGCACCCCGTCTCGATCAGTACAGCATGAGCAGAGCTTATTTCTCTTGGTTGC 1075

Qy    1070 ---taaaaaacaaactcttcaagttcacgacagccgtcaacggcggcgaccgcgccatgg 1126
              |||||||||  | |   ||   |    |||| || ||||| || || |||
Db    1076 AGGGGAAAAACAAAGAGTACGTACTCGATGCCCGGATCAAGGGTGGCGAACGTGCTATGA 1135

Qy    1127 taccgatcggcacttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgc 1186
           |  || | |  |||||  |||||  | ||| ||||||||   ||    |  |  |
Db    1136 TCATGAGCAACGAGTATGACCGCGTTTTCCCGATGGACATCTATCCGGAGTATTTGCTCA 1195

Qy    1187 gcgatttaatcgtcggcgataccgacagcgcgcaggctttgggttgcttggaattggacg 1246
           | | | || |    ||| | |||||    | |||  || ||    |  ||| |||
Db    1196 AGGCTATTATAGCATTCGACATCGACAAGATGGAGGACTTAGGCATATATGAAGTGGCTC 1255

Qy    1247 aagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgt 1306
             || ||| | ||    |||   || ||      |  |||   |||       |
Db    1256 CGGAGGACTTTGCCACTTGCGAATTTGTGGATACATCCAAGATCGAGCTGCAGCGTATCG 1315

Qy    1307 tgcgcaaagtgctggaaaccattgagaaggaa 1338
           | ||| | |   |||| |   |  | ||||||
Db    1316 TTCGCGACGGCTTGGATATGCTCTATAAGGAA 1347
```

RESULT 119
HIU20229/c
LOCUS HIU20229 4926 bp DNA linear BCT 09-FEB-1995
DEFINITION Haemophilus influenzae BOLA (bolA), glutathione reductase (gor), phosphatidylserine decarboxylase (psd), 30K protein (rpmF), genes, complete cds.
ACCESSION U20229
VERSION U20229.1 GI:644850
KEYWORDS .
SOURCE Haemophilus influenzae.
ORGANISM Haemophilus influenzae
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 4926)
AUTHORS Barcak,G.J. and Heimer,S.R.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4926)
AUTHORS Barcak,G.J.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-1995) Gerard J. Barcak, Biochemistry, University of Maryland at Baltimore, 108 North Greene Street, Baltimore, MD 21201, USA
FEATURES Location/Qualifiers
source 1. .4926

```
                /organism="Haemophilus influenzae"
                /strain="Rd"
                /isolate="KW20"
                /db_xref="taxon:727"
CDS             complement(27. .224)
                /note="orf65"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62134.1"
                /db_xref="GI:644851"
                /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                VREGDVVKKGQVLFEDKKILV"
gene            complement(519. .842)
                /gene="bolA"
CDS             complement(519. .830)
                /gene="bolA"
                /function="stationary phase regulatory protein"
                /codon_start=1
                /transl_table=11
                /product="BOLA"
                /protein_id="AAA62135.1"
                /db_xref="GI:644852"
                /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                R"
misc_feature    775. .785
                /note="encodes DNA uptake recognition sequence"
RBS             complement(838. .842)
                /gene="bolA"
protein_bind    877. .913
                /note="OxyR recongition sequence (Tartaglia, L.A., et al.,
                J. Mol Biol. 210: 709-710, 1989)"
                /bound_moiety="OxyR"
RBS             915. .919
                /note="orf199"
CDS             924. .1523
                /note="orf199"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62136.1"
                /db_xref="GI:644853"
                /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
gene            1343. .3109
                /gene="gor"
-35_signal      1343. .1348
                /gene="gor"
                /note="putative"
-10_signal      1365. .1370
                /gene="gor"
                /note="putative"
-35_signal      1585. .1590
                /gene="gor"
                /note="putative"
-10_signal      1605. .1610
                /gene="gor"
                /note="putative"
RBS             1728. .1731
                /gene="gor"
CDS             1739. .3109
                /gene="gor"
                /EC_number="1.6.4.2"
                /function="reduces oxidized glutathione using NADPH"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="glutathione reductase"
                /protein_id="AAA62137.1"
                /db_xref="GI:644854"
                /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                EFVTMR"
gene            complement(3218. .4090)
                /gene="psd"
CDS             complement(3218. .4090)
                /gene="psd"
                /function="lipid metabolism"
                /codon_start=1
                /transl_table=11
                /product="phosphatidylserine decarboxylase"
                /protein_id="AAA62138.1"
                /db_xref="GI:644855"
                /translation="MMTSNSYWQRLKVAFQYVMPQIYLTQIAGWFAKQKWGKITHFVI
                KAFAKKYNIDMSIAQKEQFSDYASFNEFFIRPLKENARPINQNPTALCLPADGRISEC
                GHIDDNLLLQAKGHFFSLEDLLAEDKELVETFKNGEFVTTYLSPRDYHRVHMPCDATL
                RKMIYVPGDLFSVNPFLAQHVPNLFARNERVICVFDTEFGTMVQILVGATITASIGTT
                WAGVINPPRHNEVKTWTYEGESAVKLLKGQEMGWFQLGSTVINLFQANQVRLADHLSV
                NEPVRMGEILAYKK"
misc_feature    complement(3364. .3374)
                /gene="psd"
                /note="encodes DNA uptake recognition sequence"
misc_feature    complement(3780. .3790)
                /gene="psd"
                /note="encodes DNA uptake recognition sequence"
misc_feature    3818. .3825
                /note="encodes DNA uptake recognition sequence"
RBS             complement(4097. .4100)
                /note="psd"
protein_bind    complement(4126. .4152)
                /note="OxyR recognition sequence (Toledano, M.B. et al.,
                Cell 78:897-909.)"
                /bound_moiety="OxyR"
gene            4247. .4771
                /gene="rpmF"
CDS             4247. .4771
                /gene="rpmF"
                /codon_start=1
                /transl_table=11
                /product="30K protein"
                /protein_id="AAA62139.1"
                /db_xref="GI:644856"
                /translation="MQKVKLPLTVDPIKDAQRRLDYVGYYAANQLERLAESVVNVLSD
                AQVTLSFYVDPQKLVVMKGKVQIDVDLECQRCNEPYKQTLECEFTYSPVANWDQADDL
                PEIYEPIEFNEFGEIDLIGTVEDELILALPLVPMHSSEHCEVSAQEQVFGVLPEELAK
                KPNPFAVLANLKQK"
CDS             complement(4561. .>4926)
                /note="orf121"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAA62140.1"
                /db_xref="GI:644857"
                /translation="LQLRGGVSEFHRLLYLLIQQQLLKHRVNDAYHDVNEIFCFVEQP
                LAILLNLIYFCFKLANTANGFGFFANSSGNTPKTCSCADTSQCSDECIGTSGKAKISS
                SSTVPIKSISPNSLNSIGS"
BASE COUNT      1598 a    862 c    938 g   1528 t
ORIGIN


  Query Match             7.8%;  Score 105.4;  DB 1;  Length 4926;
  Best Local Similarity   70.0%;  Pred. No. 9.7e-16;
  Matches  156;  Conservative    0;  Mismatches   66;  Indels    1;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| ||| || |||||| |  |||| || || ||||| | ||| |  ||||| ||
Db    224 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 165
```

Mon J

```
Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
           ||  ||||   ||  | |   |||| ||| | || || || || ||||| || |||||
Db     164 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 105

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           || || |||||  |    ||||| || |  || |||||||| ||||| ||
Db     104 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTGAAGAC 45

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggca 223
           ||| |||||  || ||| | |||||  || || ||    || |
Db      44 AAAAA-AATCCTGGTGTAATTTTTACAGCCCCTGCAAGCGGTA 3
```

RESULT 20
BACGNTZA/c

```
LOCUS       BACGNTZA               36200 bp    DNA     linear   BCT 05-FEB-1999
DEFINITION  Bacillus subtilis 36kb sequence between gntZ and trnY genes
            encoding 34 ORFs.
ACCESSION   D78193
VERSION     D78193.1  GI:1064780
KEYWORDS    phospho-bete-glucosidase; 6-phosphogluconate dehydrogenase; alkyl
            hydroperoxide reductase; amino acid permease; arginase; ornithine
            aminotransferase; positive regulatory protein; gntZ; ahp; bglA;
            yyd; yyc; formaldehyde dehydrogenase.
SOURCE      Bacillus subtilis (strain:168) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 36200)
  AUTHORS   Ogasawara,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1995) Naotake Ogasawara, Nara Institute of
            Science and Technology, Graduate School of Biological Sciences;
            8916-5 Takayama-cho, Ikoma, Nara 630-01, Japan
            (E-mail:nogasawa@bs.aist-nara.ac.jp, Tel:07437-2-5430,
            Fax:07437-2-5439)
REFERENCE   2  (sites)
  AUTHORS   Zhang,J. and Aronson,A.
  TITLE     A Bacillus subtilis bglA gene encoding phospho-beta-glucosidase is
            inducible and closely linked to a NADH dehydrogenase-encoding gene
  JOURNAL   Gene 140 (1), 85-90 (1994)
  MEDLINE   94171085
REFERENCE   3  (sites)
  AUTHORS   Hartford,O.M. and Dowds,B.C.
  TITLE     Isolation and characterization of a hydrogen peroxide resistant
            mutant of Bacillus subtilis
  JOURNAL   Microbiology 140 (Pt 2), 297-304 (1994)
  MEDLINE   94236234
REFERENCE   4  (sites)
  AUTHORS   Ogasawara,N., Nakai,S. and Yoshikawa,H.
  TITLE     Systematic sequencing of the 180 kilobase region of the Bacillus
            subtilis chromosome containing the replication origin
  JOURNAL   DNA Res. 1 (1), 1-14 (1994)
  MEDLINE   96051385
REFERENCE   5  (sites)
  AUTHORS   Calogero,S., Gardan,R., Glaser,P., Schweizer,J., Rapoport,G. and
            Debarbouille,M.
  TITLE     RocR, a novel regulatory protein controlling arginine utilization
            in Bacillus subtilis, belongs to the NtrC/NifA family of
            transcriptional activators
  JOURNAL   J. Bacteriol. 176 (5), 1234-1241 (1994)
  MEDLINE   94156824
REFERENCE   6  (sites)
  AUTHORS   Yoshida,K., Seki,S., Fujimura,M., Miwa,Y. and Fujita,Y.
  TITLE     Cloning and sequencing of a 36-kb region of the Bacillus subtilis
            genome between the gnt and iol operons
  JOURNAL   DNA Res. 2 (2), 61-69 (1995)
  MEDLINE   96093926
REFERENCE   7  (sites)
  AUTHORS   Gardan,R., Rapoport,G. and Debarbouille,M.
  TITLE     Expression of the rocDEF operon involved in arginine catabolism in
            Bacillus subtilis
  JOURNAL   J. Mol. Biol. 249 (5), 843-856 (1995)
  MEDLINE   95311309
REFERENCE   8  (bases 1 to 36200)
  AUTHORS   Kasahara,Y., Nakai,S., Yoshikawa,H. and Ogasawara,N.
  TITLE     36kb sequence between gntZ and trnY of B. subtilis genome
  JOURNAL   Unpublished (1997)
FEATURES             Location/Qualifiers
     source          1. .36200
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            1. .32
                     /gene="gntZ"
     CDS             <1. .32
                     /gene="gntZ"
                     /citation=[6]
                     /codon_start=3
                     /transl_table=11
                     /product="6-phosphogluconate dehydrogenase"
                     /protein_id="BAA11267.1"
                     /db_xref="GI:1064781"
                     /translation="GVFHTNWSE"
     gene            496. .1059
                     /gene="ahpC"
     CDS             496. .1059
                     /gene="ahpC"
                     /note="induced by hydrogen peroxide"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="alkyl hydroperoxide reductase"
                     /protein_id="BAA11268.1"
                     /db_xref="GI:1064782"
                     /translation="MSLIGKEVLPFEAKAFKNGEFIDVTNEDLKGQWSVFCFYPADFS
                     FVCPTELEDLQEQYAALKELGVEVYSVSTDTHFVHKGWHDSSEKISKITYAMIGDPSQ
                     TISRNFDVLDEETGLADRGTFIIDPDGVIQTVEINAGGIGRDASNLVNKVKAAQYVRQ
                     NPGEVCPAKWEEGGETLTPSLDLVGKI"
     gene            1073. .2602
                     /gene="ahpF"
     CDS             1073. .2602
                     /gene="ahpF"
                     /note="induced by hydrogen peroxide"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="alkyl hydroperoxide reductase"
                     /protein_id="BAA11269.1"
                     /db_xref="GI:1064783"
                     /translation="MVLDANIKAQLNQYMQLIENDIVLKVSAGEDDTSKDMLALVDEL
                     ASMSSKISVEKAELNRTPSFSVNRVGEDTGVTFAGIPLGHEFTSLVLALLQVSGRPPK
                     VDQKVIDQVKKISGEYHFESYISLTCHNCPDVVQALNMMSVLNPNITHTMIDGAAYKA
                     EVESKNIMAVPTVYLNGESFGSGRMTLEEILAKMGSGTDASEFADKEPFDVLVVGGGP
                     AGASAAIYTARKGIRTGVVAERFGGQVLDTMSIENFISVKATEGPKLAASLEEHVKEY
                     DIDVMNLQRAKRLEKKDLFELELENGAVLKSKTVILSTGARWRNVNVPGEQEFKNKGV
                     AYCPHCDGPLFEGKDVAVIGGGNSGIEAAIDLAGIVNHVTVLEFAPELKADEVLQKRL
                     YSLPNVTVVKNAQTKEITGDQSVNGITYVDRETGEEKHVELQGVFVQIGLVPNTEWLE
                     GTVERNRMGEIIVDKHGATSVPGLFAAGDCTDSAYNQIIISMGSGATAALGAFDYLIR
                     N"
     gene            complement(2712. .4151)
                     /gene="bglA"
     CDS             complement(2712. .4151)
                     /gene="bglA"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="phospho-bete-glucosidase"
                     /protein_id="BAA11270.1"
                     /db_xref="GI:1064784"
                     /translation="MGNMPKDFLWGGALAAHQFEGGWNQGGKGPSVVDVMTAGAHGVP
                     RKITDTIEENEFYPNHEAIDFYHRYKEDIALFAEMGLKCLRTSIGWSRIFPKGDEAEP
                     NEAGLQFYDDVFDELLKHGIEPVITLSHFEMPLHLAREYGGFRNRKVVDFFVNFAEAC
                     FTRYKDKVKYWMTFNEINNQMDVNNPLFLWTNSGVVVGENENAKEVMYQTAHHELVAS
                     ALAVAKGKDINPEFQIGAMVSHVPIYPFSSNPEDVMLAEEEMRQRYFFPDVQVRGYYP
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 12:42:41 ; Search time 27321.5 Seconds
(without alignments)
3375.485 Million cell updates/sec

Title: US-09-303-518D-653
Perfect score: 4407
Sequence: 1 atgaaaacaaccgacaaacg.....aattaggctaccgctggtaa 4407

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
	1	4407	100.0	4407	6	A96620	A96620 Sequence 65
	2	3954.8	89.7	4374	6	A96616	A96616 Sequence 64
	3	3954.8	89.7	4374	6	AX044007	AX044007 Sequence
c	4	3954.8	89.7	10074	1	AE002547	AE002547 Neisseria
c	5	3954.8	89.7	349980	6	AX044034	AX044034 Sequence
	6	3946.8	89.6	4374	1	NME242535	AJ242535 Neisseria
	7	3943.4	89.5	4386	1	NME296277	AJ296277 Neisseria
	8	3441.4	78.1	4350	1	NME296276	AJ296276 Neisseria
	9	3441.4	78.1	349061	1	NMA222491	AL162753 Neisseria
	10	3385.8	76.8	4350	6	A96618	A96618 Sequence 65
	11	2501.6	56.8	4355	6	A96614	A96614 Sequence 64
c	12	1327	30.1	13818	1	U32710	U32710 Haemophilus
	13	1309.2	29.7	4317	1	HIU11024	U11024 Haemophilus
	14	1297.8	29.4	4319	6	AR157320	AR157320 Sequence
	15	328.4	7.5	578	1	NLIGA3327	AJ001740 Niesseria
	16	321.6	7.3	584	1	NLIGA3354	AJ001739 Niesseria
	17	290.2	6.6	572	1	NLIGA823	AJ001737 Niesseria
	18	275.4	6.2	572	1	NLIGA3293	AJ001738 Niesseria
	19	194.8	4.4	2992	1	AF012207	AF012207 Neisseria
c	20	192.4	4.4	10485	1	AE002424	AE002424 Neisseria
c	21	192.4	4.4	349980	6	AX044030	AX044030 Sequence
	22	190	4.3	1492	1	NMIGA476	X82481 N.meningiti
	23	190	4.3	2983	1	AF012208	AF012208 Neisseria
	24	190	4.3	2995	1	AF012209	AF012209 Neisseria
	25	190	4.3	4659	1	AF235032	AF235032 Neisseria
	26	189.8	4.3	1489	1	NMIGA771	X82468 N.meningiti
	27	189.8	4.3	1489	1	NMIGATH2	X82469 N.meningiti
	28	189.8	4.3	1525	1	NMIGAG117	X82483 N.meningiti
	29	188.8	4.3	4899	1	NGIGAPRO	X04835 Neisseria g
	30	188.8	4.3	4899	6	A02796	A02796 N.gonorrhoe
	31	188.8	4.3	4899	6	A12416	A12416 IgA-proteas
	32	188.8	4.3	4899	6	AR079202	AR079202 Sequence
	33	188.8	4.3	8393	1	NGCOMLGN	Z49895 N.gonorrhoe
	34	188.4	4.3	2992	1	AF012206	AF012206 Neisseria
	35	188.4	4.3	2995	1	AF012210	AF012210 Neisseria
	36	188	4.3	1489	1	NMIGA139	X82477 N.meningiti
	37	186.4	4.2	1489	1	NMIGA341	X82478 N.meningiti
	38	186.4	4.2	1489	1	NMIGAG80	X82479 N.meningiti
	39	185.2	4.2	2980	1	AF012203	AF012203 Neisseria
	40	185.2	4.2	2980	1	AF012204	AF012204 Neisseria
	41	185.2	4.2	2980	1	AF012205	AF012205 Neisseria
c	42	185.2	4.2	311321	1	NMA3Z2491	AL162754 Neisseria
	43	185	4.2	1525	1	NMIGAC65	X82484 N.meningiti
	44	183.6	4.2	1489	1	NMIGAF48	X82480 N.meningiti
	45	183.6	4.2	1489	1	NMIGAF96	X82475 N.meningiti
	46	183.6	4.2	2980	1	AF012211	AF012211 Neisseria
	47	183.2	4.2	1525	1	NMIGAC16	X82485 N.meningiti
	48	182	4.1	1489	1	NMIGA894	X82476 N.meningiti
	49	180.8	4.1	11288	1	AE002549	AE002549 Neisseria
	50	180.8	4.1	172325	6	AX044035	AX044035 Sequence
	51	180.8	4.1	349980	6	AX044034	AX044034 Sequence
	52	180.4	4.1	5261	1	NMIGAF13	X82474 N.meningiti
	53	180.2	4.1	1489	1	NMIGA027	X82472 N.meningiti
	54	180.2	4.1	1489	1	NMIGA117	X82470 N.meningiti
	55	180.2	4.1	1489	1	NMIGA159	X82471 N.meningiti
	56	180	4.1	1522	1	NMIGAO93	X82482 N.meningiti
	57	179.4	4.1	6021	1	HEAIGA1PTC	M87491 Haemophilus
	58	178.6	4.1	1489	1	NMIGAF54	X82473 N.meningiti
	59	175.2	4.0	1525	1	NMIGA166	X82486 N.meningiti
	60	169.8	3.9	5536	1	HEAIGA1PTA	M87489 Haemophilus
	61	169.8	3.9	5657	1	HIIGA1P	X59800 Haemophilus
c	62	169.8	3.9	15212	1	U32779	U32779 Haemophilus
	63	152.8	3.5	5088	1	HEAIGA1PTB	M87490 Haemophilus
	64	150	3.4	5719	1	AF369907	AF369907 Haemophil
c	65	141.6	3.2	213494	1	SFPWR100	AL391753 Shigella
	66	140	3.2	4484	1	SFSEPAGN	Z48219 Shigella fl
	67	140	3.2	5096	1	HEAIGA1PTD	M87492 Haemophilus
	68	140	3.2	5096	1	HIGA	X64357 H.influenza
	69	135.6	3.1	4113	6	AX276281	AX276281 Sequence
	70	135.6	3.1	4394	1	SFU35656	U35656 Shigella fl
	71	135.6	3.1	5639	1	AF097644	AF097644 Escherich
c	72	135.6	3.1	48254	6	AX276220	AX276220 Sequence
c	73	135.6	3.1	51334	1	AF200692	AF200692 Shigella

	74	133.4	3.0	1390	1	HIIGA635 X82488 H.influenza
c	75	129	2.9	37710	1	ECO278144 AJ278144 Escherich
	76	123	2.8	9689	7	AF151674 AF151674 Bacteriop
	77	123	2.8	11279	7	AF151091 AF151091 Prophage
	78	122.4	2.8	4128	6	AX275950 AX275950 Sequence
c	79	122.4	2.8	7654	6	AX275945 AX275945 Sequence
	80	119.6	2.7	5208	1	AF422145 AF422145 Escherich
	81	117.4	2.7	10587	1	AF218073 AF218073 Escherich
	82	117.4	2.7	12606	1	ECAJ3631 AJ223631 Escherich
	83	110.8	2.5	1413	1	HIIGA284 X82487 H.influenza
	84	91.4	2.1	1380	1	HIDNAIGA X86103 H.influenza
	85	84.4	1.9	1477	1	HIIGAA66 X82467 H.influenza
c	86	75.8	1.7	1098	1	SF2AETX1 Z47381 S.flexneri
c	87	63.8	1.4	7218	6	I66494 I66494 Sequence 14
	88	57.6	1.3	125020	9	AF429315 AF429315 Homo sapi
c	89	53.8	1.2	125020	9	AF429315 AF429315 Homo sapi
	90	53	1.2	144	6	A12818 A12818 IgA-proteas
	91	52	1.2	2070	5	BSU43200 U43200 Boreogadus
	92	51.4	1.2	2141	9	AB070002 AB070002 Macaca fa
c	93	50.8	1.2	15776	1	AE007069 AE007069 Mycobacte
c	94	50.8	1.2	37586	1	MTCY270 Z95388 Mycobacteri
c	95	50.8	1.2	37586	6	AX191745 AX191745 Sequence
	96	49.8	1.1	2073	9	BC016843 BC016843 Homo sapi
	97	49	1.1	10732	6	E32986 E32986 Gene encodi
	98	49	1.1	313450	1	AL596170 AL596170 Listeria
c	99	48.8	1.1	142239	8	AP003266 AP003266 Oryza sat
	100	48.4	1.1	15311	1	AE007164 AE007164 Mycobacte

ALIGNMENTS

```
RESULT   1
A96620
LOCUS       A96620                  4407 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 653 from Patent WO9924578.
ACCESSION   A96620
VERSION     A96620.1  GI:6780202
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4407)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 653 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4407
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     1273 a   1169 c   1094 g    871 t
ORIGIN
```

Query Match 100.0%; Score 4407; DB 6; Length 4407;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccctaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAAAACCGGCCGC 60

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCCCCAAGCC 120

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTCGGCAAATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300

Qy     301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaatt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATT 420

Qy     421 gtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCAT 480

Qy     481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATG 540

Qy     541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGC 600

Qy     601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATT 660

Qy     661 gcaagcgcatattcttggctcgtcggtggcaatacctttgcacaaaatggatcaggtggt 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy     721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttttaccaaca 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTTACCAACA 780

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840

Qy     841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

Qy     901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960

Qy     961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGA 1020

Qy    1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTT 1080

Qy    1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140

Qy    1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttccttattgacaaa 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAA 1200

Qy    1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 1260

Qy    1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1261 GAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320

Qy    1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380

Qy    1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGC 1440

Qy    1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCC 1500

Qy    1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAG 1560

Qy    1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620

Qy    1621 tcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 TCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680

Qy    1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggcaat 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAAT 1740

Qy    1741 aacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 AACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800

Qy    1801 gcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcgcact 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 GCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACT 1860

Qy    1861 ttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaaactg 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 TTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTG 1920

Qy    1921 tttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtcaaaa 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 TTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAA 1980

Qy    1981 atggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcacattt 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 ATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTT 2040

Qy    2041 aaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgccaaa 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 AAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAA 2100

Qy    2101 gtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcaccgcat 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 GTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCAT 2160

Qy    2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccgaaaaa 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAA 2220

Qy    2221 accattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaatgtc 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTC 2280

Qy    2281 agccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaatctt 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 AGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTT 2340

Qy    2341 agtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaacctc 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 AGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTC 2400

Qy    2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460

Qy    2461 gcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtctgacg 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 GCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACG 2520

Qy    2521 ctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctcccta 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 CTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTA 2580

Qy    2581 gccgataaggcagtattccattttgaaaacagccgctttaccggaaaaatcagcggcggc 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 GCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGC 2640

Qy    2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacggaatta 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTA 2700

Qy    2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcg 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCG 2760

Qy    2761 gcaggcgcgcaaaccggcagtgcggcagatgcgccgcgccgccgttcgcgccgttcccta 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 GCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTA 2820

Qy    2821 ttatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacggtaaacggc 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 TTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGC 2880

Qy    2881 aaattgaacggtcagggaacattccgctttatgtcggaactcttcggctaccgcagcggc 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 AAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGC 2940

Qy    2941 aaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaacaataccggc 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 AAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGC 3000

Qy    3001 aacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaacacaccgctg 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 AACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTG 3060

Qy    3061 tccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgcatggcgttat 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 TCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTAT 3120

Qy    3121 cagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaagagctt 3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121 CAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3180

Qy    3181 tccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaacaggcacaa 3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181 TCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAA 3240

Qy    3241 cttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcg 3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3241 CTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCG 3300

Qy    3301 gccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccggcaggcaggc 3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3301 GCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGC 3360

Qy    3361 ggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaa 3420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3361 GGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAA 3420

Qy    3421 gacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccaccgccttcccc 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3421 GACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCC 3480
```

```
Qy   3481 cgcgcccgccgcgcccgccgggatttgccgcaaccgcagccccaaccgcaaccccaaccg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCG 3540

Qy   3541 cagcgcgacctgatcagccgttatgccaatagcggtttgagtgaattttccgccacgctc 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTC 3600

Qy   3601 aacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagaccgccgcaac 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 AACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAAC 3660

Qy   3661 gccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagatttccgcgcc 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCC 3720

Qy   3721 taccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggcagcgggcgc 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 TACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGC 3780

Qy   3781 gtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggcatcggcaac 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 GTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAAC 3840

Qy   3841 tcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcaggttcgacatc 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 TCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3900

Qy   3901 ggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatcagaggcaaa 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 GGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAA 3960

Qy   3961 atccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggtttcggcgga 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 ATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGA 4020

Qy   4021 ttcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcggattaccga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 TTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGA 4080

Qy   4081 tacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgcgcgggcatt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 TACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATT 4140

Qy   4141 aaggcagattattcattcaaaccggcgcaacacatttccatcacgccttatttgagcctg 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 AAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4200

Qy   4201 tcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgccgtattggcg 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 TCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCG 4260

Qy   4261 caggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatcaaaggtttc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 CAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTC 4320

Qy   4321 acgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcagcacagcgcg 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 ACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGCG 4380

Qy   4381 ggcatcaaattaggctaccgctggtaa 4407
          |||||||||||||||||||||||||||
Db   4381 GGCATCAAATTAGGCTACCGCTGGTAA 4407


RESULT   2
A96616
LOCUS       A96616                  4374 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 649 from Patent WO9924578.
ACCESSION   A96616
VERSION     A96616.1  GI:6780200
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4374)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 649 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     1266 a   1163 c   1066 g    879 t
ORIGIN


  Query Match             89.7%;  Score 3954.8;  DB 6;  Length 4374;
  Best Local Similarity   94.0%;  Pred. No. 0;
  Matches 4153;  Conservative    0;  Mismatches  212;  Indels   51;  Gaps    2;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy     61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
          ||||||||||||| ||||||||||||||||||||||||||||||||||| |||||||||
Db     61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy    121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
          |||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||||||
Db    241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy    301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy    361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaaatt 420
          || |||||||||||||||| ||||| |||||||||||| || || ||| ||||  |||||
Db    361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy    421 gtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcat 480
          |||||| | |||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    421 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 480

Qy    481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
          |||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||
Db    481 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 540

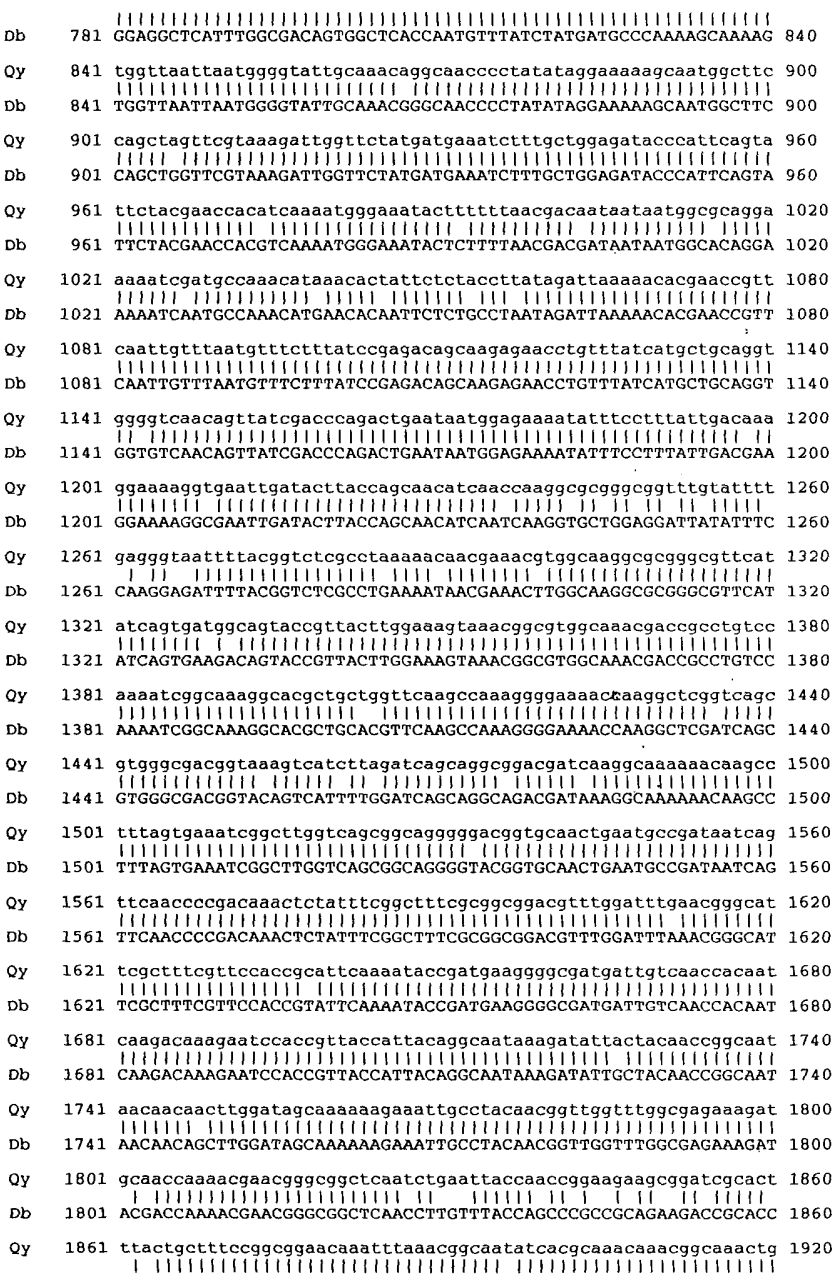
Qy    541 gatgggtggaaatacgctgatttaaataaatacccctgatcgtgttcgaatcggagcaggc 600
          |||||| |||||||     |||  |||||| |||||||| |||||||| || || ||||||
Db    541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600

Qy    601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
          || |||||||||| |||||||||| || ||||||||||||||||||||||||||||||||||
Db    601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660

Qy    661 gcaagcgcatattcttggctcgtcggtggcaataccttgcacaaaatggatcaggtggt 720
          ||||| || ||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy    721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttttaccaaca 780
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780

Qy    781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
```


Db 781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
Qy 841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
Db 841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900
Qy 901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
Db 901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960
Qy 961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
Db 961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 1020
Qy 1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
Db 1021 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080
Qy 1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db 1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140
Qy 1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttccttttattgacaaa 1200
Db 1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200
Qy 1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
Db 1201 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 1260
Qy 1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
Db 1261 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320
Qy 1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
Db 1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380
Qy 1381 aaaatcggcaaaggcacgctgctggttcaagccaaagggggaaaaccaaggctcggtcagc 1440
Db 1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440
Qy 1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
Db 1441 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500
Qy 1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
Db 1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560
Qy 1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
Db 1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1620
Qy 1621 tcgctttcgttccaccgcattcaaaatacCgatgaaggggcgatgattgtcaaccacaat 1680
Db 1621 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
Qy 1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggcaat 1740
Db 1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 1740
Qy 1741 aacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
Db 1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
Qy 1801 gcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcgcact 1860
Db 1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860
Qy 1861 ttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaaactg 1920
Db 1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
Qy 1921 tttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtcaaaa 1980
Db 1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 1980
Qy 1981 atggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcacattt 2040
Db 1981 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 2040
Qy 2041 aaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgccaaa 2100
Db 2041 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 2100
Qy 2101 gtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcaccgcat 2160
Db 2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 2160
Qy 2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccgaaaaa 2220
Db 2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 2220
Qy 2221 accattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaatgtc 2280
Db 2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 2280
Qy 2281 agccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaatctt 2340
Db 2281 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 2340
Qy 2341 agtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaacctc 2400
Db 2341 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 2400
Qy 2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
Db 2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460
Qy 2461 gcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtctgacg 2520
Db 2461 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520
Qy 2521 ctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctcccta 2580
Db 2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580
Qy 2581 gccgataaggcagtattccatttttgaaaacagccgctttaccggaaaaatcagcggcggc 2640
Db 2581 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 2640
Qy 2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacggaatta 2700
Db 2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700
Qy 2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcg 2760
Db 2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 2760
Qy 2761 gcaggcgcgcaaaccggcagtgcggcagatgcgccgcgc---------cgccgttcgcgc 2811
Db 2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 2820
Qy 2812 cgttccctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacg 2871
Db 2821 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 2880
Qy 2872 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2931
Db 2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2940
Qy 2932 cgcagcggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaac 2991
Db 2941 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000

```
Qy   2992 aataccggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaac 3051
          |||||||||||||||| | ||| ||||| ||||||||||||||||||||||||||||||||||||
Db   3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 3060

Qy   3052 acaccgctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgca 3111
          | |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3061 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 3120

Qy   3112 tggcgttatcagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3171
          |||||||| || || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3180

Qy   3172 caagagctttccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaa 3231
          |||||||||||||||||||||||||||||| | |
Db   3181 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAA---------------------------- 3213

Qy   3232 caggcacaacttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcg 3291
                    |||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   3214 ---------------GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCG 3258

Qy   3292 ctgattgcggccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccgg 3351
          |||||||||||||||||| |||||  ||||||| ||||||| ||||||||||||||||||
Db   3259 CTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318

Qy   3352 caggcaggcggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcag 3411
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3378

Qy   3412 gcggataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccacc 3471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3379 GCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACC 3438

Qy   3472 gccttcccccgcgcccgccgcgcccgcgggatttgccgcaaccgcagccccaaccgcaa 3531
          |||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||
Db   3439 GCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAG 3498

Qy   3532 ccccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaatttttcc 3591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3499 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCC 3558

Qy   3592 gccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagac 3651
          |||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||
Db   3559 GCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618

Qy   3652 cgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagat 3711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGAT 3678

Qy   3712 ttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggc 3771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3679 TTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGC 3738

Qy   3772 agcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggc 3831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3739 AGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGC 3798

Qy   3832 atcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcagg 3891
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||||
Db   3799 ATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACAGG 3858

Qy   3892 ttcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatc 3951
          ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   3859 TTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATC 3918

Qy   3952 agaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggt 4011
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||
Db   3919 GGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGT 3978

Qy   4012 ttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcg 4071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3979 TTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCG 4038

Qy   4072 gattaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgc 4131
          |||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   4039 GATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGC 4098

Qy   4132 gcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgccttat 4191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4099 GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTAT 4158

Qy   4192 ttgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgcc 4251
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db   4159 TTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCC 4218

Qy   4252 gtattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatc 4311
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4219 GTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATC 4278

Qy   4312 aaaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcag 4371
          |||||||||||||||||||||||||||||||||||||| || |||||| ||||||||||
Db   4279 AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAA 4338

Qy   4372 cacagcgcgggcatcaaattaggctaccgctggtaa 4407
          |||||||||||||||||||||||||||||||||||||
Db   4339 CACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4374


RESULT   3
AX044007
LOCUS       AX044007                4374 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 86 from Patent WO0066791.
ACCESSION   AX044007
VERSION     AX044007.1  GI:11342893
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4374)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 86 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
BASE COUNT     1266 a   1163 c   1066 g    879 t
ORIGIN


  Query Match             89.7%;  Score 3954.8;  DB 6;  Length 4374;
  Best Local Similarity   94.0%;  Pred. No. 0;
  Matches 4153;  Conservative    0;  Mismatches  212;  Indels   51;  Gaps    2;

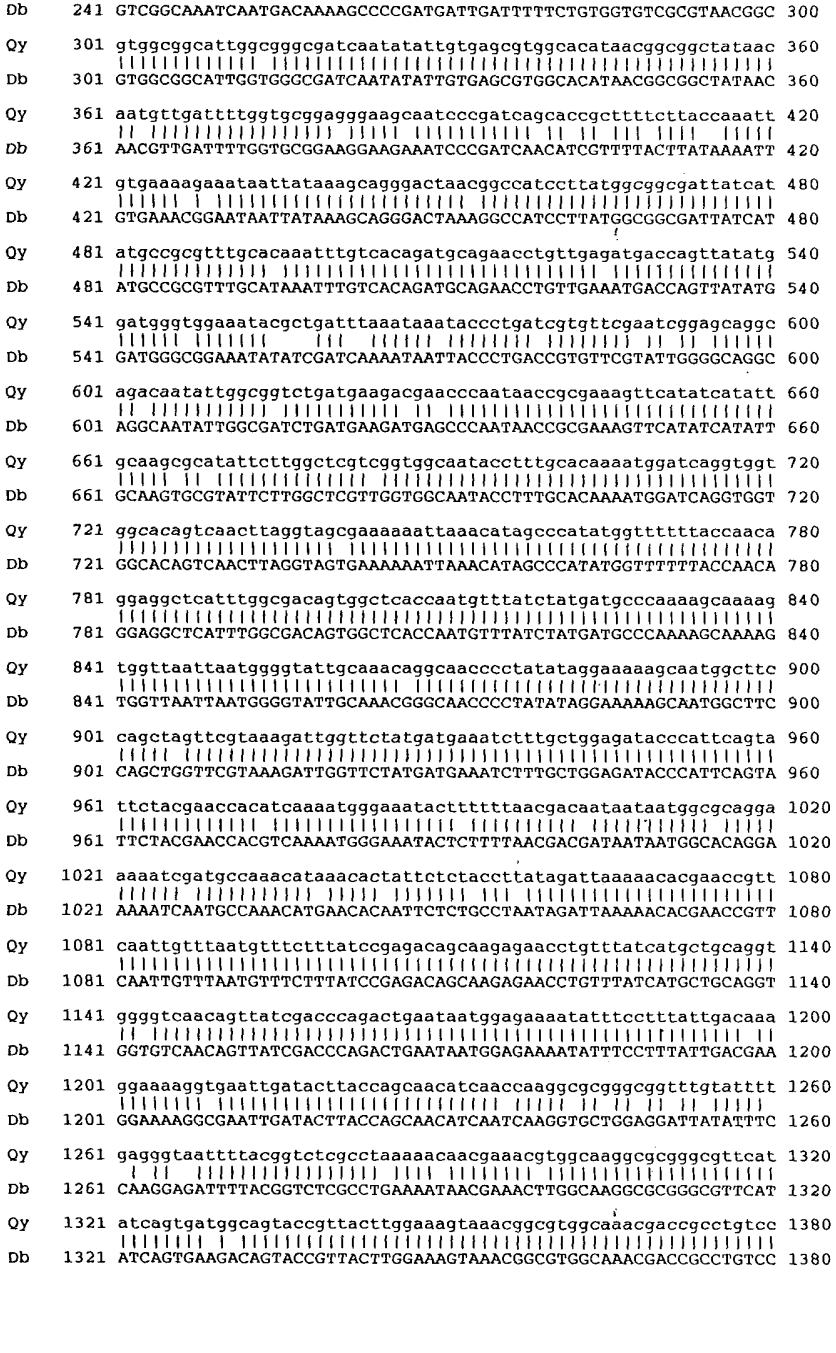
Qy      1 atgaaaacaaccgacaaaccggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy     61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
          |||||||||||||| ||||||||||||||||||||||||||||||||||||| |||||||||
Db     61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy    121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
          |||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||||||
```


Db 241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300
Qy 301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
Db 301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360
Qy 361 aatgttgatttttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaaatt 420
Db 361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420
Qy 421 gtgaaaagaaataattatataaagcagggactaacggccatccttatggcggcgattatcat 480
Db 421 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 480
Qy 481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
Db 481 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 540
Qy 541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
Db 541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600
Qy 601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
Db 601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660
Qy 661 gcaagcgcatattcttggctcgtcggtggcaatacctttgcacaaaatggatcaggtggt 720
Db 661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720
Qy 721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttaccaaca 780
Db 721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780
Qy 781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
Db 781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
Qy 841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
Db 841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900
Qy 901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
Db 901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960
Qy 961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
Db 961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 1020
Qy 1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
Db 1021 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080
Qy 1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db 1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140
Qy 1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttccttttattgacaaa 1200
Db 1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200
Qy 1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
Db 1201 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 1260
Qy 1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
Db 1261 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320
Qy 1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
Db 1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380
Qy 1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
Db 1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440
Qy 1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
Db 1441 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500
Qy 1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
Db 1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560
Qy 1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
Db 1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1620
Qy 1621 tcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
Db 1621 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
Qy 1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggcaat 1740
Db 1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 1740
Qy 1741 aacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
Db 1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
Qy 1801 gcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcgcact 1860
Db 1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860
Qy 1861 ttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaaactg 1920
Db 1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
Qy 1921 tttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtcaaaa 1980
Db 1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 1980
Qy 1981 atggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcacattt 2040
Db 1981 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 2040
Qy 2041 aaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgccaaa 2100
Db 2041 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 2100
Qy 2101 gtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcaccgcat 2160
Db 2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 2160
Qy 2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccgaaaaa 2220
Db 2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 2220
Qy 2221 accattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaatgtc 2280
Db 2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 2280
Qy 2281 agccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaatctt 2340
Db 2281 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 2340
Qy 2341 agtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaacctc 2400
Db 2341 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 2400
Qy 2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
Db 2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460

```
Qy   2461 gcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtctgacg 2520
          ||||||| |||||||||||||||||||||||| || ||||||||||||||||||||||||
Db   2461 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520

Qy   2521 ctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctcccta 2580
          ||||||| |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580

Qy   2581 gccgataaggcagtattccattttgaaaacagccgctttaccggaaaaatcagcggcggc 2640
          |||||||||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db   2581 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 2640

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacggaatta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcg 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 2760

Qy   2761 gcaggcgcgcaaaccggcagtgcggcagatgcgccgcgc---------cgccgttcgcgc 2811
          ||||| |||||||||||||||| ||||||||||||||         |||||||||||||
Db   2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 2820

Qy   2812 cgttccctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacg 2871
          |||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||||
Db   2821 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 2880

Qy   2872 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2940

Qy   2932 cgcagcggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaac 2991
          |||||| |||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2941 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000

Qy   2992 aataccggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaac 3051
          |||||||||||||||| | ||| ||||| ||||||||||||||||||||||||||||||
Db   3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 3060

Qy   3052 acaccgctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgca 3111
          | |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3061 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 3120

Qy   3112 tggcgttatcagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3171
          |||||||| || || ||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3180

Qy   3172 caagagctttccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaa 3231
          |||||||||||||||||||||||||| | |                            
Db   3181 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAA---------------------------- 3213

Qy   3232 caggcacaacttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcg 3291
                    |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3214 ----------GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCG 3258

Qy   3292 ctgattgcggccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccgg 3351
          ||||||||||||||||| |||||  ||||||| ||||||| ||||||||||||||||||
Db   3259 CTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318

Qy   3352 caggcaggcggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcag 3411
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3378

Qy   3412 gcggataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccacc 3471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3379 GCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACC 3438

Qy   3472 gccttcccccgcgcccgccgcgcccgccgggatttgccgcaaccgcagccccaaccgcaa 3531
          |||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||
Db   3439 GCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAG 3498

Qy   3532 ccccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaattttcc 3591
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3499 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCC 3558

Qy   3592 gccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagac 3651
          |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||
Db   3559 GCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618

Qy   3652 cgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagat 3711
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGAT 3678

Qy   3712 ttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggc 3771
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3679 TTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGC 3738

Qy   3772 agcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggc 3831
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   3739 AGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGC 3798

Qy   3832 atcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcagg 3891
          |||||||||||||||||||||||||||| |||||||||||||||||||||||| ||||
Db   3799 ATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACAGG 3858

Qy   3892 ttcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatc 3951
          ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   3859 TTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATC 3918

Qy   3952 agaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggt 4011
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |||
Db   3919 GGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGT 3978

Qy   4012 ttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcg 4071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3979 TTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCG 4038

Qy   4072 gattaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgc 4131
          |||||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   4039 GATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGC 4098

Qy   4132 gcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgccttat 4191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4099 GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTAT 4158

Qy   4192 ttgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgcc 4251
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db   4159 TTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCC 4218

Qy   4252 gtattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatc 4311
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4219 GTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATC 4278

Qy   4312 aaaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcag 4371
          |||||||||||||||||||||||||||||||||||||||| || |||||| ||||||||||
Db   4279 AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAA 4338

Qy   4372 cacagcgcgggcatcaaattaggctaccgctggtaa 4407
          |||||||||||||||||||||||||||||||||||
Db   4339 CACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4374
```

```
RESULT   4
AE002547/c
LOCUS       AE002547               10074 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 189 of 206
            of the complete genome.
ACCESSION   AE002547 AE002098
VERSION     AE002547.1  GI:7227244
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10074)
```

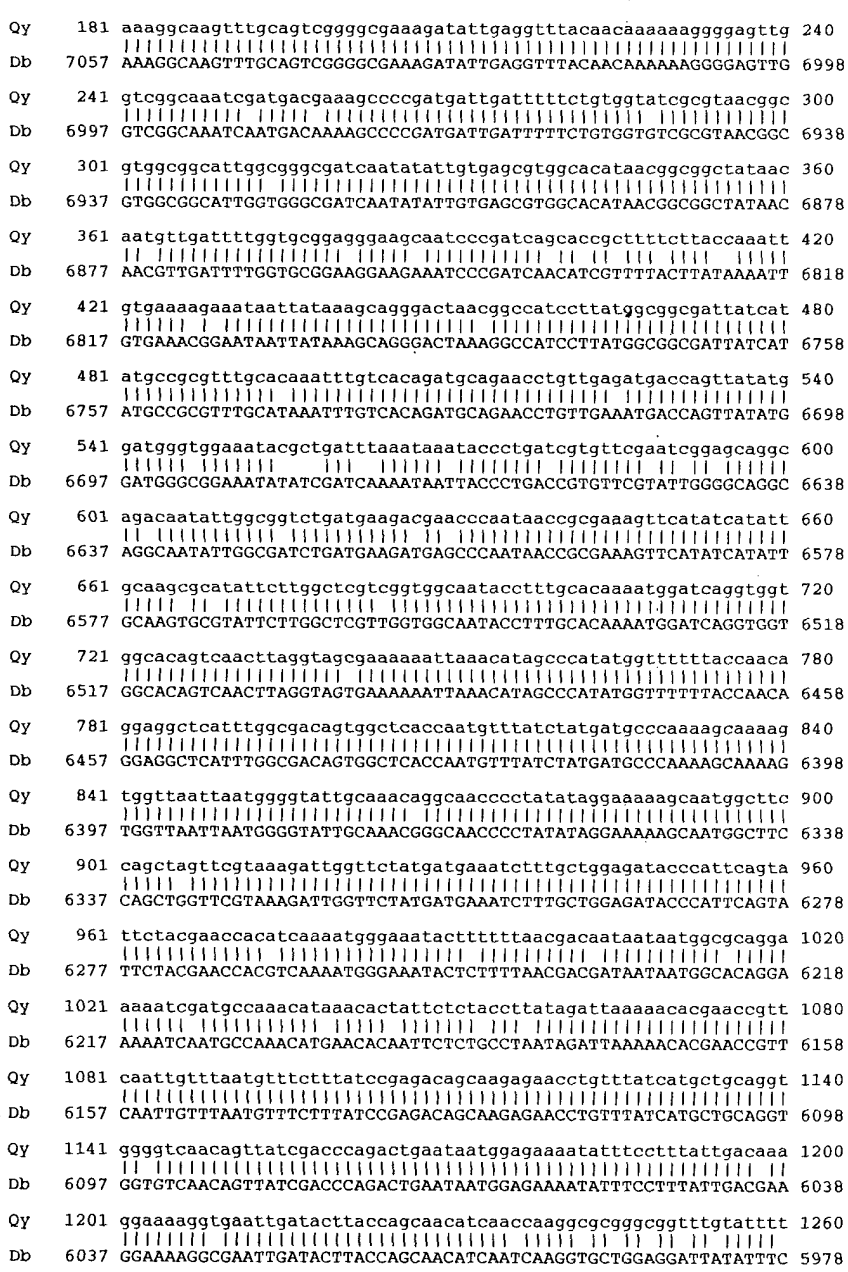
```
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10074)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10074
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            704. .1375
                     /gene="NMB1983"
     CDS             704. .1375
                     /gene="NMB1983"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF42311.1"
                     /db_xref="GI:7227245"
                     /translation="MFPDKYKLSLKENIFLAKKVLVAQIHNLSRFENCQTTLLQTEQI
                     IHGKNVASASLEDIQTILNLKRAYQYVISHISNGEPVDISLLKKINNIVAKDDSLAPG
                     DFRTGSVGVTLLDGSRHAPNPVKEIEVARVLQNIGLQSGSTTEAAVRFMLYCMRQQVF
                     WDGNKRTATLFANGLMMAGGCGILEISEMQMPQFNEKLSAFYRSGDDTDISKFVYQNC
                     ISGID"
     gene            1437. .2548
                     /gene="NMB1984"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     SP:Q00840 GB:Z11857 PID:45066 percent identity: 97.40;
                     identified by sequence similarity; putative;IS1106
                     transposase, authentic frameshift"
     gene            complement(2864. .7237)
                     /gene="NMB1985"
     CDS             complement(2864. .7237)
                     /gene="NMB1985"
                     /note="similar to GP:1408569 percent identity: 76.08;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion and penetration protein"
                     /protein_id="AAF42312.1"
                     /db_xref="GI:7227246"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVFYEPRQNGKYSFND
                     DNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINQGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNG
                     NLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHDAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFR
                     FMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFT
                     LQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALI
                     AAGRDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATT
                     AFPRARRARRDLPQLQPQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFA
                     EDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENT
                     FDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVLHYGIQ
                     ARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA
                     QHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAA
                     AAKGPQLEAQHSAGIKLGYRW"
     gene            complement(7371. .7802)
                     /gene="NMB1986"
     CDS             complement(7371. .7802)
                     /gene="NMB1986"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF42313.1"
                     /db_xref="GI:7227247"
                     /translation="MTENAQDKARQAVETVVKSPELVEQILSDEYVQIMIARCFHSGP
                     LPPPSDLAQYNDIISNGADRIMAMAEKEQAVRHETIRQDQTFNRRGQLYGFISAILIL
                     LFAVFLVWSGYPATAASLAGGTVFALAGAFVIGRSRDQGKN"
     gene            8087. .9433
                     /gene="NMB1987"
     CDS             8087. .9433
                     /gene="NMB1987"
                     /note="similar to GB:L10328 SP:P25522 PID:290554 GB:U00096
                     PID:2367268 percent identity: 69.20; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="thiophene and furan oxidation protein ThdF"
                     /protein_id="AAF42314.1"
                     /db_xref="GI:7227248"
                     /translation="MSDNVPTIAAVATAPGRGGVGVIRISGKNLLPMAQALCGKTPKP
                     RTATYADFTDTDGQAIDSGLLLFFAAPASFTGEDVIELQGHGGPVVMDMLLNRCLELG
                     ARLAEPGEFTKRAFLNDKLDLAQAEGVADLIDASSRSAARLALRSLKGDFSRRIHGLV
                     EDLITLRMLVEATLDFPEEDIDFLEAADARGKLDGLRRAVDDVLANAQQGAILREGLN
                     VVLVGAPNVGKSSLLNALAGDEVAIVTDIAGTTRDAVRERILIDGVPVHIVDTAGLRE
                     TDDVVERIGIERSRKAVSEADVALVLVDPREGLNEKTRAILDALPPELKRIEIHSKSD
                     LHAHAAGGFGTGAETVIALSAKTGDGLDALKRTLLREAGWQGESEGLFLARTRHVNAL
                     KAAQEELSLAALCGNHQIELFAEHLRLAQVACGEITGEFTADDLLGVIFSRFCIGK"
BASE COUNT     2307 a   2467 c   2721 g   2579 t
ORIGIN
```

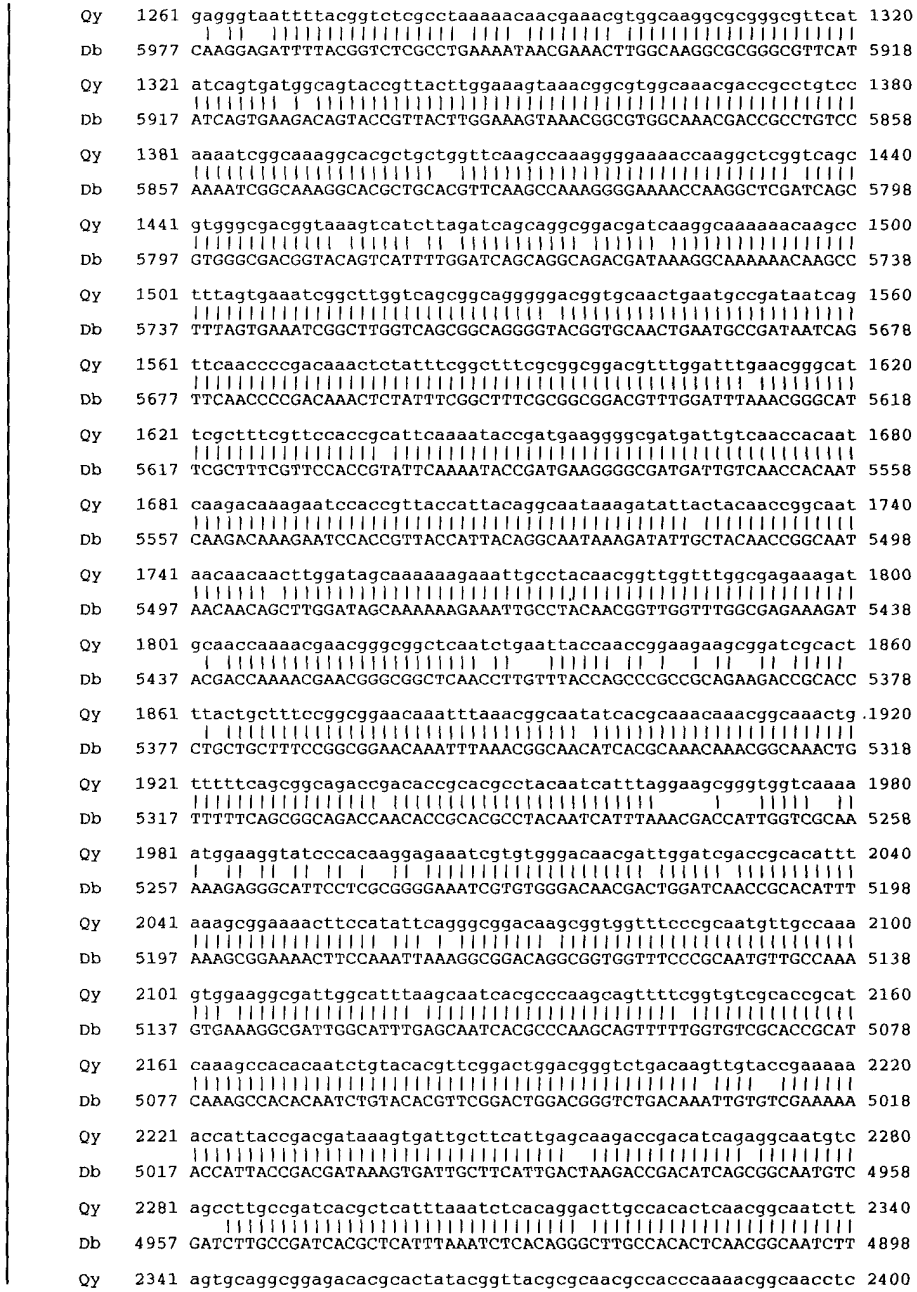
```
  Query Match             89.7%;  Score 3954.8;  DB 1;  Length 10074;
  Best Local Similarity   94.0%;  Pred. No. 0;
  Matches 4153;  Conservative    0;  Mismatches  212;  Indels   51;  Gaps    2;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    7237 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 7178

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
           |||||||||||||| ||||||||||||||||||||||||||||||||| |||||||||
Db    7177 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 7118

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7117 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 7058
```


Qy 181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
Db 7057 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 6998
Qy 241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
Db 6997 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 6938
Qy 301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
Db 6937 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 6878
Qy 361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaatt 420
Db 6877 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 6818
Qy 421 gtgaaaagaaattaattataaagcagggactaacggccatccttatggcggcgattatcat 480
Db 6817 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 6758
Qy 481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
Db 6757 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 6698
Qy 541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
Db 6697 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 6638
Qy 601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
Db 6637 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 6578
Qy 661 gcaagcgcatattcttggctcgtcggtggcaatacctttgcacaaaatggatcaggtggt 720
Db 6577 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 6518
Qy 721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttaccaaca 780
Db 6517 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 6458
Qy 781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
Db 6457 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 6398
Qy 841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
Db 6397 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 6338
Qy 901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
Db 6337 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 6278
Qy 961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
Db 6277 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 6218
Qy 1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
Db 6217 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 6158
Qy 1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db 6157 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 6098
Qy 1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttccttttattgacaaa 1200
Db 6097 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 6038
Qy 1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
Db 6037 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 5978


Qy 1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
Db 5977 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 5918
Qy 1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
Db 5917 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 5858
Qy 1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
Db 5857 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 5798
Qy 1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
Db 5797 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 5738
Qy 1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
Db 5737 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 5678
Qy 1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
Db 5677 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 5618
Qy 1621 tcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
Db 5617 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 5558
Qy 1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggcaat 1740
Db 5557 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 5498
Qy 1741 aacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
Db 5497 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 5438
Qy 1801 gcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcgcact 1860
Db 5437 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 5378
Qy 1861 ttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaaactg 1920
Db 5377 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 5318
Qy 1921 tttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtcaaaa 1980
Db 5317 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 5258
Qy 1981 atggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcacattt 2040
Db 5257 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 5198
Qy 2041 aaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgccaaa 2100
Db 5197 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 5138
Qy 2101 gtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcaccgcat 2160
Db 5137 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 5078
Qy 2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccgaaaaa 2220
Db 5077 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 5018
Qy 2221 accattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaatgtc 2280
Db 5017 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 4958
Qy 2281 agccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaatctt 2340
Db 4957 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 4898
Qy 2341 agtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaacctc 2400

```
Db   4897 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 4838

Qy   2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
Db   4837 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 4778

Qy   2461 gcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtctgacg 2520
Db   4777 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 4718

Qy   2521 ctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctcccta 2580
Db   4717 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 4658

Qy   2581 gccgataaggcagtattccattttgaaaacagccgctttaccggaaaaatcagcggcggc 2640
Db   4657 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 4598

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacggaatta 2700
Db   4597 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 4538

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcg 2760
Db   4537 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 4478

Qy   2761 gcaggcgcgcaaaccggcagtgcggcagatgcgccgcgc---------cgccgttcgcgc 2811
Db   4477 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 4418

Qy   2812 cgttccctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacg 2871
Db   4417 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 4358

Qy   2872 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2931
Db   4357 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 4298

Qy   2932 cgcagcggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaac 2991
Db   4297 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 4238

Qy   2992 aataccggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaac 3051
Db   4237 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 4178

Qy   3052 acaccgctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgca 3111
Db   4177 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 4118

Qy   3112 tggcgttatcagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3171
Db   4117 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 4058

Qy   3172 caagagctttccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaa 3231
Db   4057 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAA---------------------------- 4023

Qy   3232 caggcacaacttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcg 3291
Db   4024 ---------------GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCG 3980

Qy   3292 ctgattgcggccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccgg 3351
Db   3979 CTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3920

Qy   3352 caggcaggcggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcag 3411
Db   3919 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3860

Qy   3412 gcggataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccacc 3471
```

```
Db   3859 GCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACC 3800

Qy   3472 gccttcccccgcgcccgccgcgcccgccgggatttgccgcaaccgcagccccaaccgcaa 3531
Db   3799 GCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAG 3740

Qy   3532 ccccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaattttcc 3591
Db   3739 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCC 3680

Qy   3592 gccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagac 3651
Db   3679 GCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3620

Qy   3652 cgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagat 3711
Db   3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGAT 3560

Qy   3712 ttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggc 3771
Db   3559 TTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGC 3500

Qy   3772 agcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggc 3831
Db   3499 AGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGC 3440

Qy   3832 atcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcagg 3891
Db   3439 ATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACAGG 3380

Qy   3892 ttcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatc 3951
Db   3379 TTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATC 3320

Qy   3952 agaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggt 4011
Db   3319 GGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGT 3260

Qy   4012 ttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcg 4071
Db   3259 TTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCG 3200

Qy   4072 gattaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgc 4131
Db   3199 GATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGC 3140

Qy   4132 gcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgccttat 4191
Db   3139 GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTAT 3080

Qy   4192 ttgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgcc 4251
Db   3079 TTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCC 3020

Qy   4252 gtattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatc 4311
Db   3019 GTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATC 2960

Qy   4312 aaaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcag 4371
Db   2959 AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAA 2900

Qy   4372 cacagcgcgggcatcaaattaggctaccgctggtaa 4407
Db   2899 CACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 2864
```

```
RESULT   5
AX044034/c
LOCUS       AX044034              349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 113 from Patent WO0066791.
ACCESSION   AX044034
VERSION     AX044034.1  GI:11342918
```

```
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 113 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.-seq 1: 1 to
                     349980 349980 bases-seq 108: 300001 to 649980 349980
                     bases-seq 109: 600001 to 949980 349980 bases-seq 110:
                     900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
                     349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
                     113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    86473 a  95646 c  85908 g  81953 t
ORIGIN
```

Query Match 89.7%; Score 3954.8; DB 6; Length 349980;
Best Local Similarity 94.0%; Pred. No. 0;
Matches 4153; Conservative 0; Mismatches 212; Indels 51; Gaps 2;

```
Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db 289191 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 289132

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
          ||||||||||||| |||||||||||||||||||||||||||||||||| |||||||||||
Db 289131 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 289072

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 289071 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 289012

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 289011 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 288952

Qy     241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
           |||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||||||
Db 288951 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 288892

Qy     301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db 288891 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 288832

Qy     361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaatt 420
           || |||||||||||||||| ||||| |||||||||||| || || ||| |||| |||||
Db 288831 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 288772

Qy     421 gtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcat 480
           |||||| | ||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 288771 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 288712

Qy     481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
           |||||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||
Db 288711 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 288652

Qy     541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
           |||||| |||||||    ||| |||||| |||||||| |||||||| || || ||||||
Db 288651 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 288592

Qy     601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
           || |||||||||||| |||||||||||| || ||||||||||||||||||||||||||||
```

```
Db 288591 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 288532

Qy     661 gcaagcgcatattcttggctcgtcggtggcaataccttgcacaaaatggatcaggtggt 720
           ||||| || |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db 288531 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 288472

Qy     721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttttaccaaca 780
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db 288471 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 288412

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288411 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 288352

Qy     841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db 288351 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 288292

Qy     901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288291 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 288232

Qy     961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
           |||||||||||| |||||||||||||||||| |||||||||| |||||||||| |||||
Db 288231 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 288172

Qy    1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
           |||||| |||||||||| ||||| ||||||| ||| ||||||||||||||||||||||||
Db 288171 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 288112

Qy    1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288111 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 288052

Qy    1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttccttttattgacaaa 1200
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 288051 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 287992

Qy    1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
           |||||||| |||||||||||||||||||||||||| ||||| || || || || |||||
Db 287991 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 287932

Qy    1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
           | ||  |||||||||||||||| |||| |||||||| ||||||||||||||||||||||
Db 287931 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 287872

Qy    1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
           |||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||
Db 287871 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 287812

Qy    1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
           |||||||||||||||||||  |||||||||||||||||||||||||||| |||||
Db 287811 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 287752

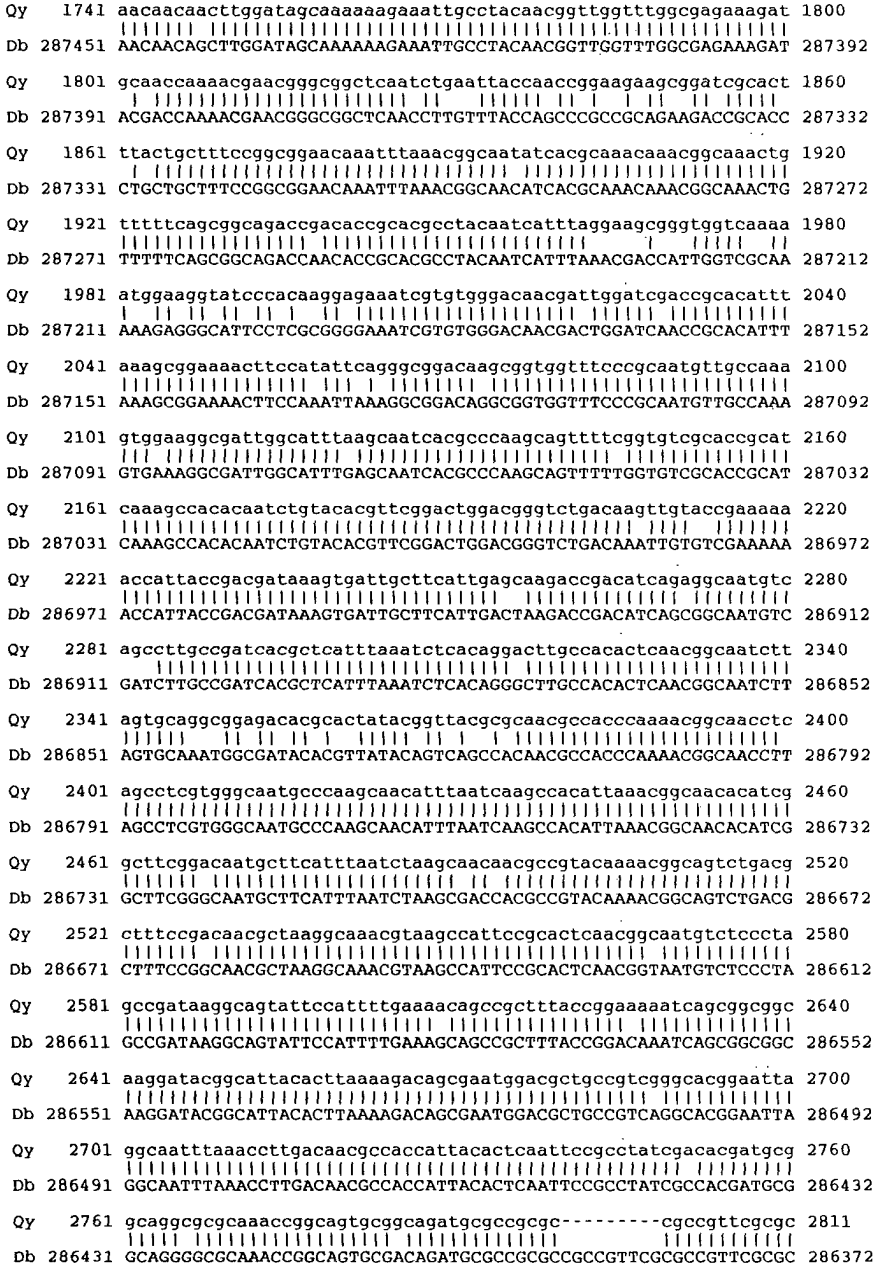
Qy    1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
           |||||||||||| |||||| || |||||||||| |||||| ||||||||||||||||||
Db 287751 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 287692

Qy    1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 287691 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 287632

Qy    1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db 287631 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 287572

Qy    1621 tcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db 287571 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 287512

Qy    1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggcaat 1740
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db 287511 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 287452
```


Qy 1741 aacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
Db 287451 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 287392
Qy 1801 gcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcgcact 1860
Db 287391 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 287332
Qy 1861 ttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaaactg 1920
Db 287331 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 287272
Qy 1921 tttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtcaaaa 1980
Db 287271 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 287212
Qy 1981 atggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcacattt 2040
Db 287211 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 287152
Qy 2041 aaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgccaaa 2100
Db 287151 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 287092
Qy 2101 gtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcaccgcat 2160
Db 287091 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 287032
Qy 2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccgaaaaa 2220
Db 287031 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 286972
Qy 2221 accattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaatgtc 2280
Db 286971 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 286912
Qy 2281 agccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaatctt 2340
Db 286911 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 286852
Qy 2341 agtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaacctc 2400
Db 286851 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 286792
Qy 2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
Db 286791 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 286732
Qy 2461 gcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtctgacg 2520
Db 286731 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 286672
Qy 2521 ctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctcccta 2580
Db 286671 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 286612
Qy 2581 gccgataaggcagtattccatttgaaaacagccgctttaccggaaaaatcagcggcggc 2640
Db 286611 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 286552
Qy 2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacggaatta 2700
Db 286551 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 286492
Qy 2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcg 2760
Db 286491 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 286432
Qy 2761 gcaggcgcgcaaaccggcagtgcggcagatgcgccgcgc---------cgccgttcgcgc 2811
Db 286431 GCAGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 286372
Qy 2812 cgttccctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacg 2871
Db 286371 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 286312
Qy 2872 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2931
Db 286311 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 286252
Qy 2932 cgcagcggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaac 2991
Db 286251 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 286192
Qy 2992 aataccggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaac 3051
Db 286191 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 286132
Qy 3052 acaccgctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgca 3111
Db 286131 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 286072
Qy 3112 tggcgttatcagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3171
Db 286071 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 286012
Qy 3172 caagagctttccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaa 3231
Db 286011 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAA---------- 285977
Qy 3232 caggcacaacttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcg 3291
Db 285978 ----------GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCG 285934
Qy 3292 ctgattgcggccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccgg 3351
Db 285933 CTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 285874
Qy 3352 caggcaggcggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcag 3411
Db 285873 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 285814
Qy 3412 gcggataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccacc 3471
Db 285813 GCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACC 285754
Qy 3472 gccttcccccgcgcccgccgcgcccgcgggatttgccgcaaccgcagccccaaccgcaa 3531
Db 285753 GCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAG 285694
Qy 3532 ccccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaattttcc 3591
Db 285693 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCC 285634
Qy 3592 gccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagac 3651
Db 285633 GCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 285574
Qy 3652 cgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagat 3711
Db 285573 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGAT 285514
Qy 3712 ttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggc 3771
Db 285513 TTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGC 285454
Qy 3772 agcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggc 3831
Db 285453 AGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGC 285394
Qy 3832 atcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcagg 3891
Db 285393 ATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACAGG 285334
Qy 3892 ttcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatc 3951

```
          ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db 285333 TTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATC 285274

Qy   3952 agaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggt 4011
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||
Db 285273 GGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGT 285214

Qy   4012 ttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcg 4071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285213 TTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCG 285154

Qy   4072 gattaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgc 4131
          |||||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 285153 GATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGC 285094

Qy   4132 gcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgccttat 4191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285093 GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTAT 285034

Qy   4192 ttgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgcc 4251
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db 285033 TTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCC 284974

Qy   4252 gtattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatc 4311
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 284973 GTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATC 284914

Qy   4312 aaaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcag 4371
          |||||||||||||||||||||||||||||||||||| || |||||| ||||||||||||
Db 284913 AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAA 284854

Qy   4372 cacagcgcgggcatcaaattaggctaccgctggtaa 4407
          |||||||||||||||||||||||||||||||||||
Db 284853 CACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 284818
```

```
RESULT   6
NME242535
LOCUS       NME242535               4374 bp    DNA     linear   BCT 16-JAN-2002
DEFINITION  Neisseria meningitidis serogroup B app gene, strain B:15:P1.16.
ACCESSION   AJ242535
VERSION     AJ242535.2  GI:6977940
KEYWORDS    APP gene; app protein.
SOURCE      Neisseria meningitidis serogroup B.
  ORGANISM  Neisseria meningitidis serogroup B
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (sites)
  AUTHORS   Ala Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1999) Ala Aldeen D.A., Microbiology, University
            of Nottingham, Division of Microbiology, SCLS, University Hospital,
            Nottingham, NG7 2UH, UNITED KINGDOM
  REMARK    revised by {3}
REFERENCE   3  (bases 1 to 4374)
  AUTHORS   Ala Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) Ala Aldeen D.A., Microbiology, University
            of Nottingham, Division of Microbiology, SCLS, University Hospital,
            Nottingham, NG7 2UH, UNITED KINGDOM
COMMENT     On Feb 14, 2000 this sequence version replaced gi:4938471.
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="Neisseria meningitidis serogroup B"
                     /strain="B:15:P1.16"
                     /isolate="SD"
                     /db_xref="taxon:491"
     gene            1. .4374
                     /gene="app"
     CDS             1. .4374
                     /gene="app"
                     /codon_start=1
                     /transl_table=11
                     /product="App protein"
                     /protein_id="CAB43832.2"
                     /db_xref="GI:6977941"
                     /db_xref="SPTREMBL:Q9X7H1"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDAHSVFYEPRQNGKYSFND
                     DNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINQGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGSLSANGDTRYTVSHYATQNG
                     NLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHNAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFR
                     FMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFT
                     LQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALI
                     AAGRDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATT
                     AFPRARRARRDLPQLQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFA
                     EDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENT
                     FDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVLHYGIQ
                     ARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA
                     QHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAA
                     AAKGPQLEAQHSAGIKLGYRW"
BASE COUNT     1265 a   1163 c   1066 g    880 t
ORIGIN
```

```
  Query Match             89.6%;  Score 3946.8;  DB 1;  Length 4374;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 4148;  Conservative    0;  Mismatches   217;  Indels    51;  Gaps    2;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
           |||||||||||||| ||||||||||||||||||||||||||||||||||||| |||||||||
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

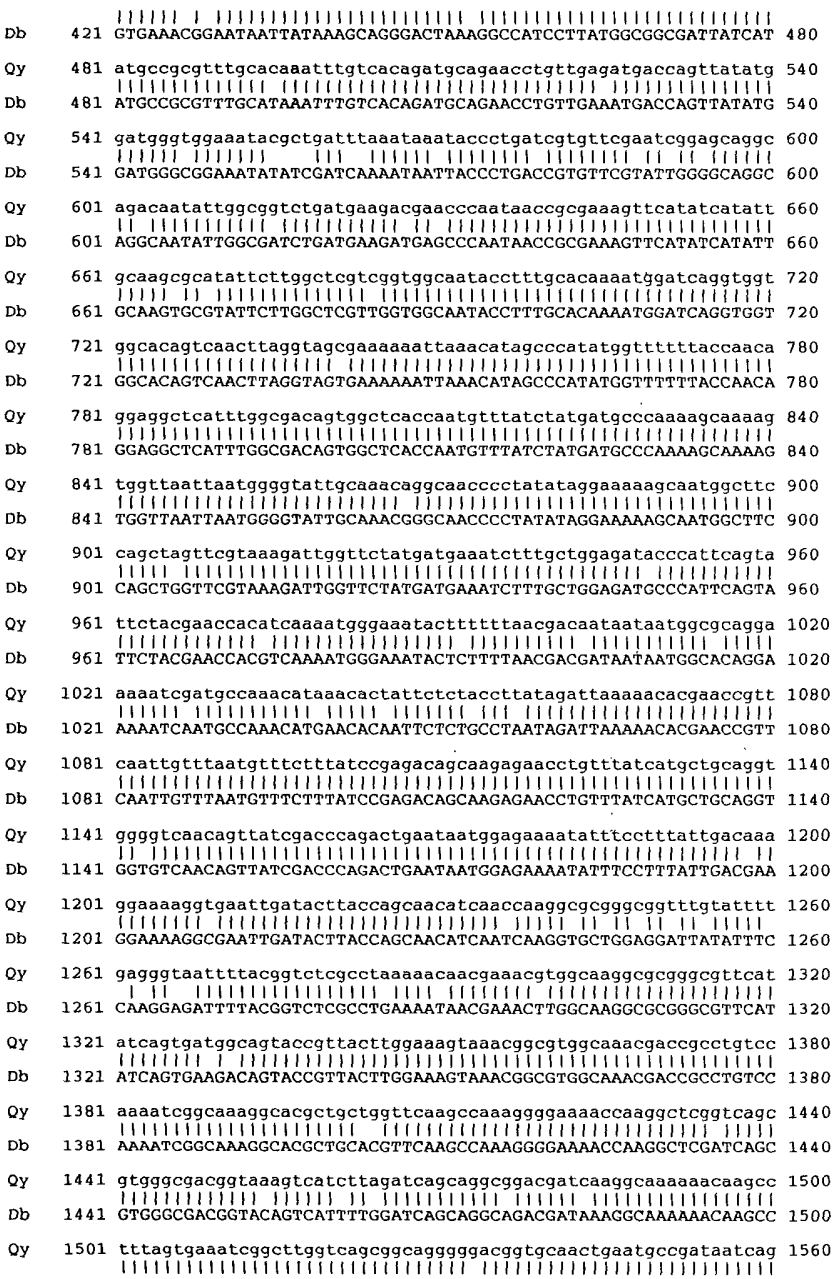
Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
           |||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaatt 420
           || |||||||||||||||| ||||| |||||||||||| || || ||| ||||  |||||
Db     361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy     421 gtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcat 480
```


Db 421 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 480
Qy 481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
Db 481 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 540
Qy 541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
Db 541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600
Qy 601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
Db 601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660
Qy 661 gcaagcgcatattcttggctcgtcggtggcaatacctttgcacaaaatggatcaggtggt 720
Db 661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720
Qy 721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttaccaaca 780
Db 721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780
Qy 781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
Db 781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
Qy 841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
Db 841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900
Qy 901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
Db 901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATGCCCATTCAGTA 960
Qy 961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
Db 961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 1020
Qy 1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
Db 1021 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080
Qy 1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db 1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140
Qy 1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacaaa 1200
Db 1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200
Qy 1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
Db 1201 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 1260
Qy 1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
Db 1261 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320
Qy 1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
Db 1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380
Qy 1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
Db 1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440
Qy 1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
Db 1441 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500
Qy 1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
Db 1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560
Qy 1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
Db 1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1620
Qy 1621 tcgctttcgttccaccgcattcaaaatacgatgaaggggcgatgattgtcaaccacaat 1680
Db 1621 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
Qy 1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggcaat 1740
Db 1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 1740
Qy 1741 aacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
Db 1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
Qy 1801 gcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcgcact 1860
Db 1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860
Qy 1861 ttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaaactg 1920
Db 1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
Qy 1921 tttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtcaaaa 1980
Db 1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 1980
Qy 1981 atggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcacattt 2040
Db 1981 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 2040
Qy 2041 aaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgccaaa 2100
Db 2041 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 2100
Qy 2101 gtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcaccgcat 2160
Db 2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 2160
Qy 2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccgaaaaa 2220
Db 2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 2220
Qy 2221 accattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaatgtc 2280
Db 2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 2280
Qy 2281 agccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaatctt 2340
Db 2281 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAGTCTT 2340
Qy 2341 agtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaacctc 2400
Db 2341 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACTACGCCACCCAAAACGGCAACCTT 2400
Qy 2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
Db 2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460
Qy 2461 gcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtctgacg 2520
Db 2461 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520
Qy 2521 ctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctcccta 2580
Db 2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580
Qy 2581 gccgataaggcagtattccatttgaaaacagccgctttaccggaaaaatcagcggcggc 2640
Db 2581 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 2640

```
Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacggaatta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||||
Db   2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCATCAGGCACGGAATTA 2700

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcg 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||||
Db   2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACAATGCG 2760

Qy   2761 gcaggcgcgcaaaccggcagtgcggcagatgcgccgcgc---------cgccgttcgcgc 2811
          ||||| |||||||||||||||||| ||||||||||||||         ||||||||||||
Db   2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 2820

Qy   2812 cgttccctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacg 2871
          |||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||||
Db   2821 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 2880

Qy   2872 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2940

Qy   2932 cgcagcggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaac 2991
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2941 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000

Qy   2992 aataccggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaac 3051
          |||||||||||||||||| | ||| |||| ||||||||||||||||||||||||||||||||
Db   3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 3060

Qy   3052 acaccgctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgca 3111
          | |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3061 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 3120

Qy   3112 tggcgttatcagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3171
          |||||||| || || ||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3180

Qy   3172 caagagctttccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaa 3231
          |||||||||||||||||||||||||||||| | |
Db   3181 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAA---------------------------- 3213

Qy   3232 caggcacaacttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcg 3291
                    |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3214 ---------------GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCG 3258

Qy   3292 ctgattgcggccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccgg 3351
          |||||||||||||||| ||||| ||||||| ||||||| ||||||||||||||||||||||
Db   3259 CTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318

Qy   3352 caggcaggcggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcag 3411
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3378

Qy   3412 gcggataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccacc 3471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3379 GCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACC 3438

Qy   3472 gccttcccccgcgcccgccgcgcccgccgggatttgccgcaaccgcagccccaaccgcaa 3531
          |||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db   3439 GCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAG 3498

Qy   3532 ccccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaatttttcc 3591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3499 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTTCC 3558

Qy   3592 gccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagac 3651
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db   3559 GCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618

Qy   3652 cgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagat 3711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGAT 3678
```

```
Qy   3712 ttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggc 3771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3679 TTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGC 3738

Qy   3772 agcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggc 3831
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   3739 AGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGC 3798

Qy   3832 atcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcagg 3891
          |||||||||||||||||||||||||||||| |||||||||||||||||||||| ||||
Db   3799 ATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACAGG 3858

Qy   3892 ttcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatc 3951
          ||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   3859 TTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATC 3918

Qy   3952 agaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggt 4011
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||
Db   3919 GGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGT 3978

Qy   4012 ttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcg 4071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3979 TTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCG 4038

Qy   4072 gattaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgc 4131
          |||||||| |||||||||||||||||||||||| ||||||||||||||||||||||||
Db   4039 GATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGC 4098

Qy   4132 gcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgccttat 4191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4099 GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTAT 4158

Qy   4192 ttgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgcc 4251
          |||||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db   4159 TTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCC 4218

Qy   4252 gtattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatc 4311
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4219 GTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATC 4278

Qy   4312 aaaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcag 4371
          |||||||||||||||||||||||||||||||||||||||||||| || |||||| ||||||||||
Db   4279 AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAA 4338

Qy   4372 cacagcgcgggcatcaaattaggctaccgctggtaa 4407
          |||||||||||||||||||||||||||||||||||
Db   4339 CACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4374
```

```
RESULT   7
NME296277
LOCUS       NME296277               4386 bp    DNA     linear   BCT 14-JAN-2002
DEFINITION  Neisseria meningitidis app gene for adhesion penetration protein,
            strain Z4181.
ACCESSION   AJ296277
VERSION     AJ296277.1  GI:11071864
KEYWORDS    adhesion penetration protein; APP gene.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (bases 1 to 4386)
  AUTHORS   Ala-Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2000) Ala-Aldeen D.A., Microbiology, University
            of Nottingham, Molecular Bacteriology and Immunology Group,
            University Hospital, Nottingham NG7 2UH, UNITED KINGDOM
```

```
FEATURES             Location/Qualifiers
     source          1. .4386
                     /organism="Neisseria meningitidis"
                     /strain="Z4181"
                     /db_xref="taxon:487"
                     /country="Mali"
                     /note="serogroup C"
     gene            1. .4365
                     /gene="app"
     CDS             1. .4365
                     /gene="app"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion penetration protein"
                     /protein_id="CAC14671.1"
                     /db_xref="GI:11071865"
                     /db_xref="SPTREMBL:Q9F3Z4"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFTYKIVKRNNYKAGTNGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVFYEPRQNGKYSFHD
                     NNNGTGKIDAKHEHNSLPNRLKTRTVQLFNVSLSETAQEPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINRGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGKVILDQQADENNKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTNTDISGNVSLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNG
                     DLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHSESSRFTGQISGSKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHDAAGAQTGSATDAPRRRSRRSLLSVTPPASVESRFNTLTVNGKLNGQGTFRFMS
                     ELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFTLQN
                     EHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALIAAG
                     RDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATTAFP
                     RARRARRDLPQLQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEGR
                     RNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDD
                     GIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGSKIRRRVLHYGIQARY
                     RAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
                     SITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAK
                     GPQLEAQHSAGIKLGYRW"
BASE COUNT      1270 a    1167 c    1066 g     883 t
ORIGIN
```

```
  Query Match             89.5%;  Score 3943.4;  DB 1;  Length 4386;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 4134;  Conservative    0;  Mismatches  231;  Indels   42;  Gaps    1;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
           ||||||||||||||| ||||||||||||||||||||||||||||||||| ||||||||
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
           |||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaatt 420
           || |||||||||||||||| |||||||||||||||| || || ||| |||| |||||
Db     361 AACGTTGATTTTGGTGCGGAAGGAAGCAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy     421 gtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcat 480
           |||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GTGAAACGGAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCAT 480

Qy     481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
           |||||||||||| || |||||||| || |||||||||||||||| ||||||||||||||
Db     481 ATGCCGCGTTTACATAAATTTGTAACCGATGCAGAACCTGTTGAAATGACCAGTTATATG 540

Qy     541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
           |||||| |||||||     |||  |||||| |||||||| |||||||| || || ||||||
Db     541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600

Qy     601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
           || |||||||||| |||||||||| || ||||||||||||||||||||||||||||||||
Db     601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660

Qy     661 gcaagcgcatattcttggctcgtcggtggcaataccttgcacaaaatggatcaggtggt 720
           ||||| || ||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy     721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttttaccaaca 780
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTTACCAACA 780

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAA 840

Qy     841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

Qy     901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960

Qy     961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
           |||||||||||| |||||||||||||||||| |||| ||||||||||||||||| |||||
Db     961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTCACGACAATAATAATGGCACAGGA 1020

Qy    1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
           |||||||| |||||||| ||||| |||||| ||| ||||||||||||||||||||||||
Db    1021 AAAATCGACGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080

Qy    1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
           |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db    1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCACAAGAACCTGTTTATCATGCTGCAGGT 1140

Qy    1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttccttattgacaaa 1200
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200

Qy    1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
           |||||||| |||||||||||||||| |||||||| ||| || || || || || |||||
Db    1201 GGAAAAGGCGAATTGATACTTACCAGTAACATCAACCGAGGTGCTGGAGGATTATATTTC 1260

Qy    1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
           | ||  |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    1261 CAAGGAGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320

Qy    1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
           |||||||| | |||||||||||||||||||| || ||||||||||||||||||||||||
Db    1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTCAATGGCGTGGCAAACGACCGCCTGTCC 1380

Qy    1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
           |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||||
Db    1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440
```


Qy 1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
Db 1441 GTGGGCGACGGTAAAGTTATTTTAGATCAACAAGCAGATGAAAATAATAAAAAACAAGCC 1500
Qy 1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
Db 1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560
Qy 1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
Db 1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
Qy 1621 tcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
Db 1621 TCACTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
Qy 1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggcaat 1740
Db 1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 1740
Qy 1741 aacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
Db 1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
Qy 1801 gcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcgcact 1860
Db 1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCTGCAGAAGACCGCACC 1860
Qy 1861 ttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaaactg 1920
Db 1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
Qy 1921 tttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtcaaaa 1980
Db 1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 1980
Qy 1981 atggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcacattt 2040
Db 1981 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 2040
Qy 2041 aaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgccaaa 2100
Db 2041 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 2100
Qy 2101 gtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcaccgcat 2160
Db 2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTGGTGTCGCACCGCAT 2160
Qy 2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccgaaaaa 2220
Db 2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 2220
Qy 2221 accattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaatgtc 2280
Db 2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGACCAATACCGACATCAGCGGCAATGTC 2280
Qy 2281 agccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaatctt 2340
Db 2281 AGCCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 2340
Qy 2341 agtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaacctc 2400
Db 2341 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCGACCTT 2400
Qy 2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
Db 2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460
Qy 2461 gcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtctgacg 2520
Db 2461 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520
Qy 2521 ctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctcccta 2580
Db 2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580
Qy 2581 gccgataaggcagtattccattttgaaaacagccgctttaccggaaaaatcagcggcggc 2640
Db 2581 GCCGATAAGGCAGTATTCCATTCTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCAGC 2640
Qy 2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacggaatta 2700
Db 2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700
Qy 2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcg 2760
Db 2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 2760
Qy 2761 gcaggcgcgcaaaccggcagtgcggcagatgcgccgcgccgccgttcgcgccgttcccta 2820
Db 2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTA 2820
Qy 2821 ttatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacggtaaacggc 2880
Db 2821 TTATCCGTTACACCTCCGGCTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGC 2880
Qy 2881 aaattgaacggtcagggaacattccgctttatgtcggaactcttcggctaccgcagcggc 2940
Db 2881 AAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGAC 2940
Qy 2941 aaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaacaataccggc 3000
Db 2941 AAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGC 3000
Qy 3001 aacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaacacaccgctg 3060
Db 3001 AACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTG 3060
Qy 3061 tccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgcatggcgttat 3120
Db 3061 TCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTAC 3120
Qy 3121 cagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaagagctt 3180
Db 3121 CAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3180
Qy 3181 tccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaacaggcacaa 3240
Db 3181 TCCGACAAACTCGGCAAGGCAGAA---------- 3204
Qy 3241 cttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcg 3300
Db 3205 ------GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCG 3258
Qy 3301 gccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccggcaggcaggc 3360
Db 3259 GCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGC 3318
Qy 3361 ggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaa 3420
Db 3319 GGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAA 3378
Qy 3421 gacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccaccgccttcccc 3480
Db 3379 GACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCC 3438
Qy 3481 cgcgcccgccgcgcccgccgggatttgccgcaaccgcagccccaaccgcaaccccaaccg 3540
Db 3439 CGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCG 3498
Qy 3541 cagcgcgacctgatcagccgttatgccaatagcggtttgagtgaattttccgccacgctc 3600
Db 3499 CAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTC 3558
Qy 3601 aacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagaccgccgcaac 3660

```
Db    3559 AACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGGCCGCCGCAAC 3618

Qy    3661 gccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagatttccgcgcc 3720
Db    3619 GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCC 3678

Qy    3721 taccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggcagcgggcgc 3780
Db    3679 TACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGC 3738

Qy    3781 gtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggcatcggcaac 3840
Db    3739 GTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAAC 3798

Qy    3841 tcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcaggttcgacatc 3900
Db    3799 TCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACATC 3858

Qy    3901 ggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatcagaggcaaa 3960
Db    3859 GGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGTCTTTCAGACGGCATCGGAAGCAAA 3918

Qy    3961 atccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggtttcggcgga 4020
Db    3919 ATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGA 3978

Qy    4021 ttcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcggattaccga 4080
Db    3979 TTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGC 4038

Qy    4081 tacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgcgcgggcatt 4140
Db    4039 TACGAAAACGTCAATATCGCCACCCCCGGCCTTGCGTTCAACCGCTACCGCGCGGGCATT 4098

Qy    4141 aaggcagattattcattcaaaccggcgcaacacatttccatcacgccttatttgagcctg 4200
Db    4099 AAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4158

Qy    4201 tcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgccgtattggcg 4260
Db    4159 TCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCT 4218

Qy    4261 caggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatcaaaggtttc 4320
Db    4219 CAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTC 4278

Qy    4321 acgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcagcacagcgcg 4380
Db    4279 ACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCG 4338

Qy    4381 ggcatcaaattaggctaccgctggtaa 4407
Db    4339 GGCATCAAATTAGGCTACCGCTGGTAA 4365


RESULT   8
NME296276
LOCUS       NME296276               4350 bp    DNA     linear   BCT 14-JAN-2002
DEFINITION  Neisseria meningitidis app gene for adhesion penetration protein,
            strain Z3515.
ACCESSION   AJ296276
VERSION     AJ296276.1  GI:11071862
KEYWORDS    adhesion penetration protein; APP gene.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (bases 1 to 4350)
  AUTHORS   Ala-Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2000) Ala-Aldeen D.A., Microbiology, University
            of Nottingham, Molecular Bacteriology and Immunology Group,
            University Hospital, Nottingham NG7 2UH, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .4350
                     /organism="Neisseria meningitidis"
                     /strain="Z3515"
                     /db_xref="taxon:487"
                     /country="Saudi Arabia"
                     /note="serogroup A"
     gene            1. .4350
                     /gene="app"
     CDS             1. .4350
                     /gene="app"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion penetration protein"
                     /protein_id="CAC14670.1"
                     /db_xref="GI:11071863"
                     /db_xref="SPTREMBL:Q9F3Z5"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHM
                     PRLHKFVTDAEPVEMTSDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGA
                     WLIGGNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLL
                     NGVLQTGYPYSGRENGFQLIRKDWFYDDIYRGDTHTVFFEPRSNGHFSFTSNNNGTGT
                     VTETNEKVSNPKLEVQTVRLFDESLNETDKEPVYAAGGVNQYRPRLNNGENLSFIDYG
                     NGKLILSNNINQGAGGLYFEGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRL
                     SKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVSGRGTVQLNA
                     DNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNATTTSTVTITGNESI
                     TQPSGKNINRLNYSKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLSGGTNLNG
                     NITQTNGKLFFSGRPTPHAYNHLGSGWSKMEGIPQGEIVWDNDWINRTFKAENFHIQG
                     GQAVISRNVAKVEGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEKTITDDKV
                     IASLTKTDISGNVSLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNGNLSLVGN
                     AQATFNQATLNGNTSASGNASFNLSNNAAQNGSLTLSDNAKANVSHSALNGNVSLADK
                     AVFHFENSRFTGQLSGSKDTALHLKDSEWTLPSGTELGNLNLDNATITLNSAYRHDAA
                     GAQTGSVSDTPRRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFRFMSELFGYRS
                     DKLKLAESSEGTYTLAVNNTGNEPVSLDQLTVVEGKDNKPLSENLNFTLQNEHVDAGA
                     WRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALIAAGRDAAEKT
                     ESVAEPARQAGGENVGIMQAEEEKKRVQADKDSALAKQREAETRPATTAFPRARRARR
                     DLPQPQPQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVW
                     TSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNS
                     ARLAHGAVFGQYGIGRFDIGISTGAGFSSGSLSDGIGGKIRRRVLHYGIQARYRAGFG
                     GFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPY
                     LSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGPQLE
                     AQHSAGIKLGYRW"
BASE COUNT     1236 a    1185 c    1073 g     856 t
ORIGIN

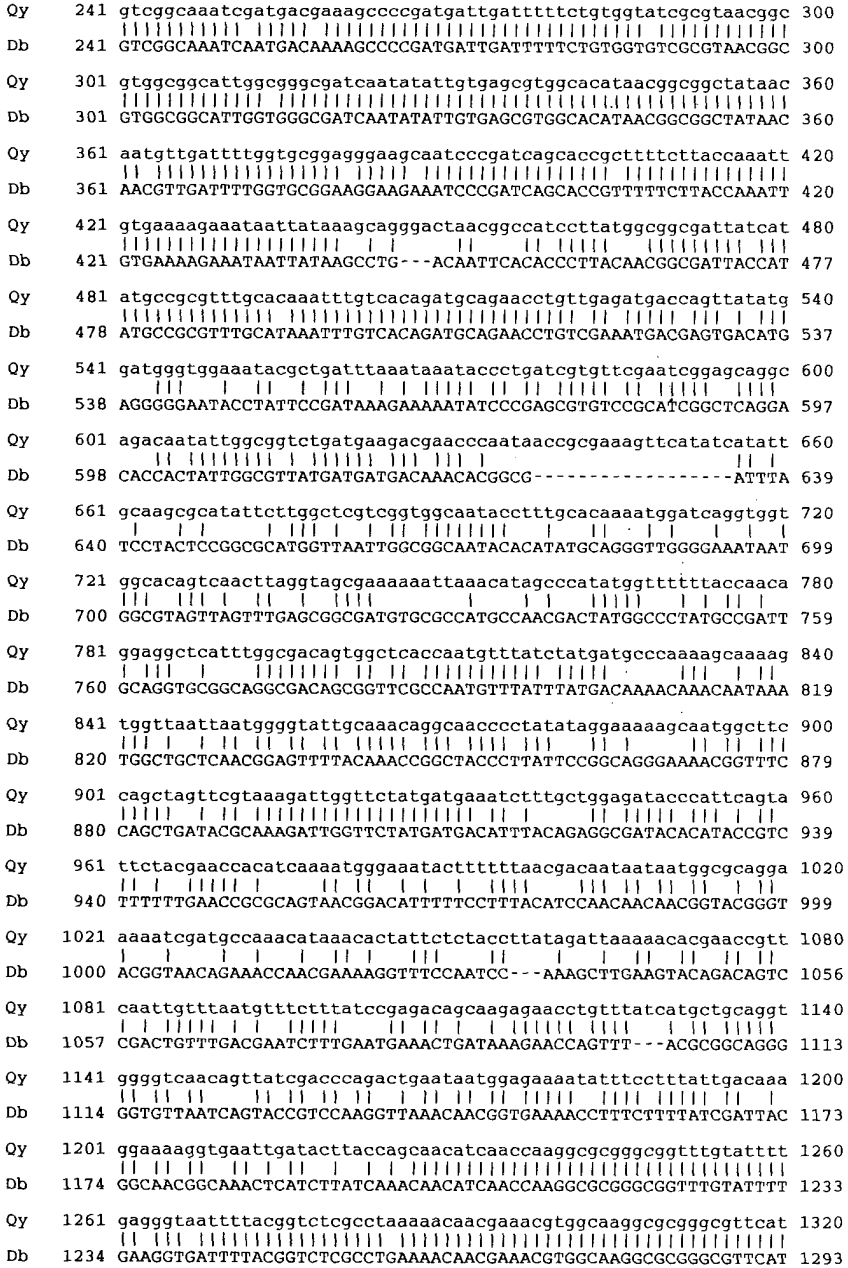

  Query Match             78.1%;  Score 3441.4;  DB 1;  Length 4350;
  Best Local Similarity   87.6%;  Pred. No. 0;
  Matches 3872;  Conservative    0;  Mismatches  466;  Indels   81;  Gaps    7;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240
```


Qy 241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
Db 241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300
Qy 301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
Db 301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360
Qy 361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaatt 420
Db 361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAGCACCGTTTTTCTTACCAAATT 420
Qy 421 gtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcat 480
Db 421 GTGAAAAGAAATAATTATAAGCCTG---ACAATTCACACCCTTACAACGGCGATTACCAT 477
Qy 481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
Db 478 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATG 537
Qy 541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
Db 538 AGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGA 597
Qy 601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
Db 598 CACCACTATTGGCGTTATGATGATGACAAACACGGCG----------ATTTA 639
Qy 661 gcaagcgcatattcttggctcgtcggtggcaataccttgcacaaaatggatcaggtggt 720
Db 640 TCCTACTCCGGCGCATGGTTAATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAAT 699
Qy 721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggttttttaccaaca 780
Db 700 GGCGTAGTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATT 759
Qy 781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
Db 760 GCAGGTGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAA 819
Qy 841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
Db 820 TGGCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTC 879
Qy 901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
Db 880 CAGCTGATACGCAAAGATTGGTTCTATGATGACATTTACAGAGGCGATACACATACCGTC 939
Qy 961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
Db 940 TTTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACAACGGTACGGGT 999
Qy 1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
Db 1000 ACGGTAACAGAAACCAACGAAAAGGTTTCCAATCC---AAAGCTTGAAGTACAGACAGTC 1056
Qy 1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db 1057 CGACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTT---ACGCGGCAGGG 1113
Qy 1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacaaa 1200
Db 1114 GGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGATTAC 1173
Qy 1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
Db 1174 GGCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 1233
Qy 1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
Db 1234 GAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1293
Qy 1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
Db 1294 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1353
Qy 1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
Db 1354 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1413
Qy 1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
Db 1414 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1473
Qy 1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
Db 1474 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1533
Qy 1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
Db 1534 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1593
Qy 1621 tcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
Db 1594 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAATCATAAT 1653
Qy 1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggc--- 1737
Db 1654 GCCACAACAACATCCACCGTTACCATTACAGGGAATGAAAGTATTACACAACCGAGTGGT 1713
Qy 1738 ---aataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
Db 1714 AAGAATATCAATAGACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1773
Qy 1795 aaagatgcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggat 1854
Db 1774 AAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGAC 1833
Qy 1855 cgcactttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggc 1914
Db 1834 CGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGC 1893
Qy 1915 aaactgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtgg 1974
Db 1894 AAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGG 1953
Qy 1975 tcaaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgc 2034
Db 1954 TCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCAACCGC 2013
Qy 2035 acatttaaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgtt 2094
Db 2014 ACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTT 2073
Qy 2095 gccaaagtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgca 2154
Db 2074 GCCAAAGTGGAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTGGTGTCGCA 2133
Qy 2155 ccgcatcaaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtacc 2214
Db 2134 CCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 2193
Qy 2215 gaaaaaaccattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggc 2274
Db 2194 GAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGC 2253
Qy 2275 aatgtcagccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggc 2334
Db 2254 AATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 2313
Qy 2335 aatcttagtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggc 2394
Db 2314 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGC 2373
Qy 2395 aacctcagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2454

```
Db   2374 AACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 2433

Qy   2455 acatcggcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagt 2514
Db   2434 ACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACAAAACGGCAGT 2493

Qy   2515 ctgacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtc 2574
Db   2494 CTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTC 2553

Qy   2575 tccctagccgataaggcagtattccattttgaaaacagccgctttaccggaaaaatcagc 2634
Db   2554 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCAGC 2613

Qy   2635 ggcggcaaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacg 2694
Db   2614 GGCAGCAAGGATACAGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACG 2673

Qy   2695 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacac 2754
Db   2674 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2733

Qy   2755 gatgcggcaggcgcgcaaaccggcagtgcggcagatgcgccgcgccgccgttcgcgccgt 2814
Db   2734 GATGCTGCAGGCGCGCAAACCGGCAGTGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGT 2793

Qy   2815 tccctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacggta 2874
Db   2794 TCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTA 2853

Qy   2875 aacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctaccgc 2934
Db   2854 AACGGCAAATTGAACGGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGA 2913

Qy   2935 agcggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaacaat 2994
Db   2914 AGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT 2973

Qy   2995 accggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaacaca 3054
Db   2974 ACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAA 3033

Qy   3055 ccgctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgcatgg 3114
Db   3034 CCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGG 3093

Qy   3115 cgttatcagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaa 3174
Db   3094 CGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA 3153

Qy   3175 gagctttccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaacag 3234
Db   3154 GAGCTTTCCGACAAACTCGGCAAGGCAGAA---------------------------- 3183

Qy   3235 gcacaacttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcgctg 3294
Db   3184 ------------GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTG 3231

Qy   3295 attgcggccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccggcag 3354
Db   3232 ATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAG 3291

Qy   3355 gcaggcggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcaggcg 3414
Db   3292 GCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCG 3351

Qy   3415 gataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccaccgcc 3474
Db   3352 GATAAAGACAGCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCC 3411

Qy   3475 ttcccccgcgcccgccgcgcccgccgggatttgccgcaaccgcagccccaaccg------ 3528
```

```
Db   3412 TTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCT 3471

Qy   3529 caaccccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaattt 3588
Db   3472 CAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTT 3531

Qy   3589 tccgccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaa 3648
Db   3532 TCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAA 3591

Qy   3649 gaccgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaa 3708
Db   3592 GACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAA 3651

Qy   3709 gatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctc 3768
Db   3652 GATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC 3711

Qy   3769 ggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgac 3828
Db   3712 GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGAC 3771

Qy   3829 ggcatcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggc 3888
Db   3772 GGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGC 3831

Qy   3889 aggttcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggc 3948
Db   3832 AGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCAGTCTTTCAGACGGC 3891

Qy   3949 atcagaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgca 4008
Db   3892 ATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCC 3951

Qy   4009 ggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaa 4068
Db   3952 GGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAA 4011

Qy   4069 gcggattaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctac 4128
Db   4012 GCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGCTAC 4071

Qy   4129 cgcgcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgcct 4188
Db   4072 CGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCT 4131

Qy   4189 tatttgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaatacc 4248
Db   4132 TATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACC 4191

Qy   4249 gccgtattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaa 4308
Db   4192 GCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAA 4251

Qy   4309 atcaaaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcg 4368
Db   4252 ATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCG 4311

Qy   4369 cagcacagcgcgggcatcaaattaggctaccgctggtaa 4407
Db   4312 CAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4350


RESULT   9
NMA2Z2491
LOCUS       NMA2Z2491              349061 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 2/7.
ACCESSION   AL162753 AL157959
VERSION     AL162753.2  GI:7379120
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
```

```
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349061)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 349061)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .349061
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     repeat_unit     complement(24. .206)
                     /note="ATR repeat; hmmfs hit to HMM ATR (1 - 183), score:
                     310.39"
                     /label=ATR
     RBS             209. .212
     gene            220. .681
                     /gene="NMA0368"
     CDS             220. .681
                     /gene="NMA0368"
                     /note="NMA0368, probable integral membrane protein, len:
                     153 aa; contains four probable transmembrane domains"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAB83669.1"
                     /db_xref="GI:7379121"
                     /db_xref="SPTREMBL:Q9JWH7"
                     /translation="MQEQNRKSSFPIVMLLVSVALWIASLSNVAFYLGNHGSMEGLTV
                     LILGSIFASLDIRYCAVYANYVWLAAIVLLALRKKVVPVHAAFWGLALVAFSVKAVYV
                     DEAGNTSDIVRYGAGFYLWYAAFAVASIGTFAGKNKERKAASAADGTKNDV"
     gene            671. .1492
                     /gene="hemK"
     CDS             671. .1492
                     /gene="hemK"
                     /note="NMA0369, hemK, HemK protein, len: 273 aa; similar
                     to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
                     scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
                     PS00092 N-6 Adenine-specific DNA methylases signature"
                     /codon_start=1
                     /transl_table=11
                     /product="HemK protein"
                     /protein_id="CAB83670.1"
                     /db_xref="GI:7379122"
                     /db_xref="SPTREMBL:Q9JWH6"
                     /translation="MTFDKWLGLSKLPKNEARMLLQYVSEYTRVQLLTRGGEEMPDEV
                     RQRADRLAQRRLNGEPVAYILGAREFYGRRFTVNPSVLIPRPETEHLVEAVLARLPEN
                     GRVWDLGTGSGAVAVTVALERPDAFVRASDISPPALETARKNAADLGARVEFAYGSWF
                     DTDMPSEGKWDIIVSNPPYIENGDKHLSQGDLRFEPQIALTDFSDGLSCIRTLAQGAP
                     DRLAEGGFLLLEHGFDQGAAVRGVLAENGFSGVETLPDLAGLDRVTLGKYMKHLK"
     misc_feature    831. .840
                     /gene="hemK"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    1162. .1171
                     /gene="hemK"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    1187. .1207
                     /gene="hemK"
                     /note="PS00092 N-6 Adenine-specific DNA methylases
                     signature"
     misc_feature    complement(1279. .1288)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             1565. .1568
     gene            1578. .2966
                     /gene="NMA0370"
     CDS             1578. .2966
                     /gene="NMA0370"
                     /note="NMA0370, probable integral membrane protein, len:
                     462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
                     P44640 hypothetical protein HI0325 (450 aa), fasta scores;
                     E(): 0, 49.8% identity in 464 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAB83671.1"
                     /db_xref="GI:7379123"
                     /db_xref="SPTREMBL:Q9JWH5"
                     /translation="MNAVVVAVIVMLVLSLSRVHVVLSLTIGAFVGGAVAGMPLQNIA
                     DAAGQVSQAGIIPVFNKGLEGGAKIALSYAMLGAFAMAITHSGLPQQLAGAVVRKLNR
                     GGMPDSVRSGEGVVKWLLLSIILVMGMMSQNIIPIHIAFIPMIVPPLLLVFNRLKIDR
                     RLIACVITFGLVTTYMFLPYGFGAIFLNEILLGNIHSAAPQLDVKNINVMAAMAIPAL
                     GMLAGLLLAFVHYRKPRLYQSNNADTAGNADAANRPQPSAYRSLAAAVAIAVCFAIQL
                     MYEDSLVLGAMLGFAVFMMLGVINRDKANDVFGEGIKMMAMVGFIMIAAQGFAAVMNA
                     TGHIQPLVESSMAIFGNSKGMAALAMLVVGLLVTMGIGSSFSTLPIIAAIYVPLCVGL
                     GFSPLATVAIVGTAGALGDAGSPASDSTLGPTMGLNADGQHDHIRDSVIPTFIHYNIP
                     LLIAGWIAAMVL"
     gene            2967. .3191
                     /gene="slyX"
     CDS             2967. .3191
                     /gene="slyX"
                     /note="NMA0371, slyX, SLYX protein homolog, len: 74 aa;
                     simlar to SLYX_HAEIN P44759 SLYX protein homolog (73 aa),
                     fasta scores; E(): 0.0017, 33.8% identity in 74 aa
                     overlap, and SLYX_ECOLI P30857 SLYX protein (72 aa), fasta
                     scores; E(): 0.24, 32.4% identity in 68 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="SLYX protein homolog"
                     /protein_id="CAB83672.1"
                     /db_xref="GI:7379124"
                     /db_xref="SPTREMBL:Q9JWH4"
                     /translation="MDAVQEFEHRITELEIQSALQEDVIAGLNAMVAELRQTLDLQQA
                     QLRLLYQKMQDRNPDAQEPYSLRDEIPPHY"
     gene            complement(3271. .3423)
                     /gene="NMA0372"
     CDS             complement(3271. .3423)
                     /gene="NMA0372"
                     /note="NMA0372, unknown, questionable CDS, len: 50 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="very hypothetical protein NMA0372"
                     /protein_id="CAB83673.1"
                     /db_xref="GI:7379125"
                     /db_xref="SPTREMBL:Q9JWH3"
                     /translation="MFGRLSPLGSDGFFLAVYEYPVGFICLQGKTAYFRSGGKRFHRI
                     KGILSD"
     misc_feature    3562. .3571
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(3586. .4356)
                     /gene="thiF"
     CDS             complement(3586. .4356)
                     /gene="thiF"
                     /note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
                     similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
```

```
                     fasta scores; E(): 0, 43.1% identity in 246 aa overlap,
                     and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
                     protein. (249 aa), fasta scores; E(): 0, 43.9% identity in
                     244 aa overlap (note that N.m. does not have orthologs of
                     any other molybdopterin biosynthesis proteins). Contains
                     Pfam match to entry PF00899 ThiF_family, ThiF family"
                     /codon_start=1
                     /transl_table=11
                     /product="ThiF protein"
                     /protein_id="CAB83674.1"
                     /db_xref="GI:7379126"
                     /db_xref="SPTREMBL:Q9JWH2"
                     /translation="MTTTEHDNDDAFLLRYSRHILLDEIGIEGQQKLSAAHILVVGCG
                     GLGAAALPYLAASGIGTLTIADSDTVELHNLQRQVAFDEGDVGKLKTEALADRLRHIN
                     HTVDVRTINEKLDGCRLTGLVQAADIVLDCCDNYATRQAVNRACVQAKTPLVSGAAVR
                     FEGQLAVYRPDLPDSPCYACLFDGGSASDGICSLFGVFSPLVGIIGSTQAAEALKILL
                     DAGEPSHGRLAVYRALEGGWQYFDLPRNPECPVCGAER"
     misc_feature    3787. .3796
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(3847. .4257)
                     /gene="thiF"
                     /note="Pfam match to entry PF00899 ThiF_family, ThiF
                     family, score 186.60, E-value 4.1e-52"
     misc_feature    complement(4418. .4427)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            4471. .7224
                     /gene="ppc"
     CDS             4471. .7224
                     /gene="ppc"
                     /EC_number="4.1.1.31"
                     /note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
                     917 aa; simlar to many e.g. CAPP_RHOPA O32483
                     phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
                     fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
                     Contains 2x Pfam match to entry PF00311 PEPcase,
                     Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
                     site motif A (P-loop), and PS00393 Phosphoenolpyruvate
                     carboxylase active site 2"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoenolpyruvate carboxylase"
                     /protein_id="CAB83675.1"
                     /db_xref="GI:7379127"
                     /db_xref="SPTREMBL:Q9JWH1"

  Query Match            78.1%;  Score 3441.4;  DB 1;  Length 349061;
  Best Local Similarity  87.6%;  Pred. No. 0;
  Matches 3872;  Conservative    0;  Mismatches  466;  Indels   81;  Gaps    7;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccctaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   94066 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 94125

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
           |||||||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||
Db   94126 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 94185

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94186 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 94245

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94246 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 94305

Qy     241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
           |||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||||||
Db   94306 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 94365

Qy     301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   94366 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 94425

Qy     361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaatt 420
           || ||||||||||||||||| ||||| ||||||||||||||||| ||||||||||||||||
Db   94426 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAGCACCGTTTTTCTTACCAAATT 94485

Qy     421 gtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcat 480
           |||||||||||||||||||  | |      ||    || |||||   |||||||| |||
Db   94486 GTGAAAAGAAATAATTATAAGCCTG---ACAATTCACACCCTTACAACGGCGATTACCAT 94542

Qy     481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
           |||||||||||||| ||||||||||||||||||||||||||| || ||||| ||| | |||
Db   94543 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATG 94602

Qy     541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
           |||  |  || | |||  | | ||||| || || ||||| || |||||  ||||
Db   94603 AGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGA 94662

Qy     601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
             || |||||||| | |||||| ||| ||| |                    || |
Db   94663 CACCACTATTGGCGTTATGATGATGACAAACACGGCG-----------------ATTTA 94704

Qy     661 gcaagcgcatattcttggctcgtcggtggcaataccttgcacaaaatggatcaggtggt 720
             |  | |    | ||| |  | || ||||||||    |   ||   | |   |  | |
Db   94705 TCCTACTCCGGCGCATGGTTAATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAAT 94764

Qy     721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttaccaaca 780
           |||  ||| |  ||  |  ||||       |    | |   |||||   | | || |
Db   94765 GGCGTAGTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATT 94824

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
           | ||| |    |||||||| || || ||||||||||  |||||      |||  | ||
Db   94825 GCAGGTGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAA 94884

Qy     841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
           ||| |  | || || || || ||||| ||| |||| |||   || |     || || |||
Db   94885 TGGCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTC 94944

Qy     901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
           |||||  | || |||||||||||||| ||||| || |     || ||||| ||| | ||
Db   94945 CAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACATACCGTC 95004

Qy     961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
           || |  ||||| |     || ||  | |  | ||||    ||| || || ||  | ||
Db   95005 TTTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACAACGGTACGGGT 95064

Qy    1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
           |  |    | |   | || |    ||     |    |  | ||| ||  || ||
Db   95065 ACGGTAACAGAAACCAACGAAAAGGTTTCCAATCC---AAAGCTTAAAGTACAGACAGTC 95121

Qy    1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
           | | ||||| | |  |||||    || || |  | |||||| ||||   | || |||||
Db   95122 CGACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTT---ACGCGGCAGGG 95178

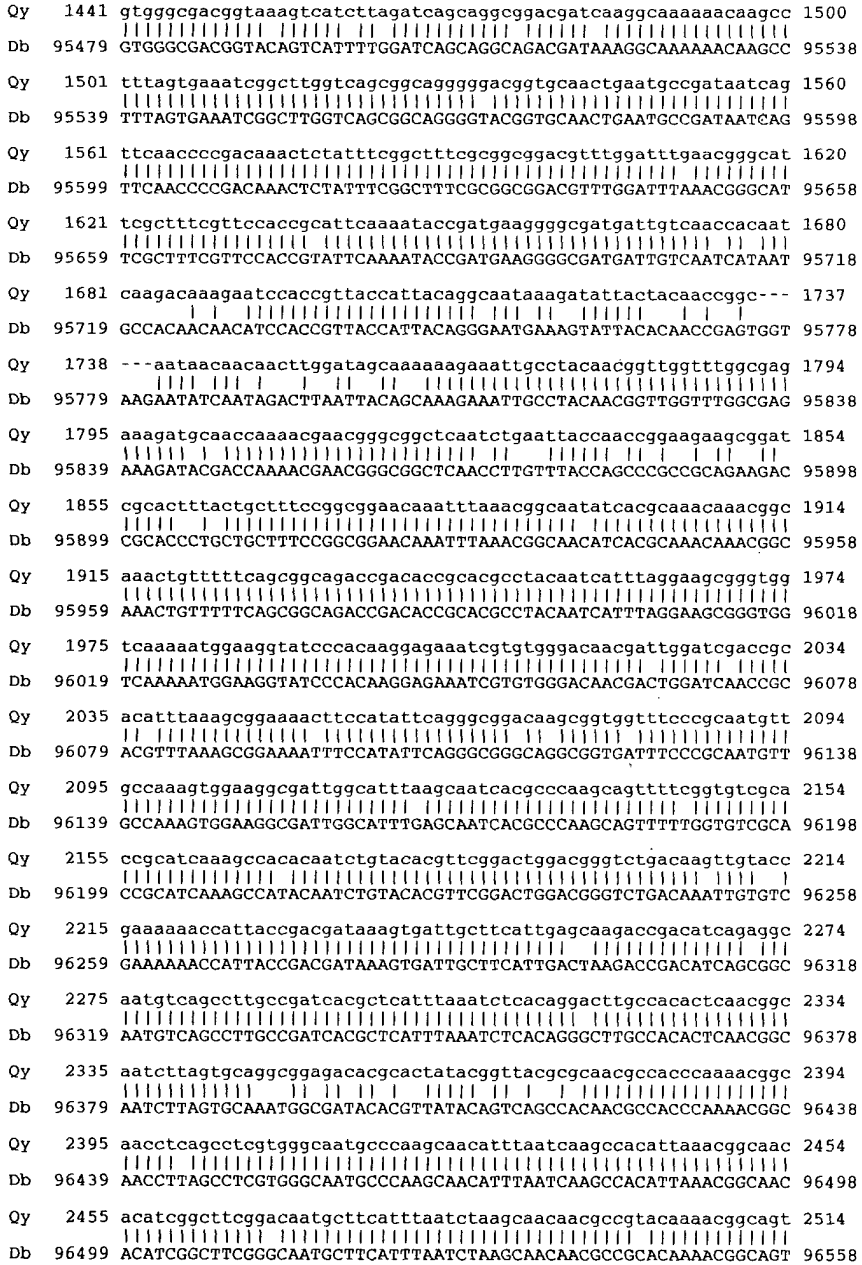
Qy    1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacaaa 1200
           || || ||    || || || ||  | || || || |||||  |||| ||||| ||  |
Db   95179 GGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGATTAC 95238

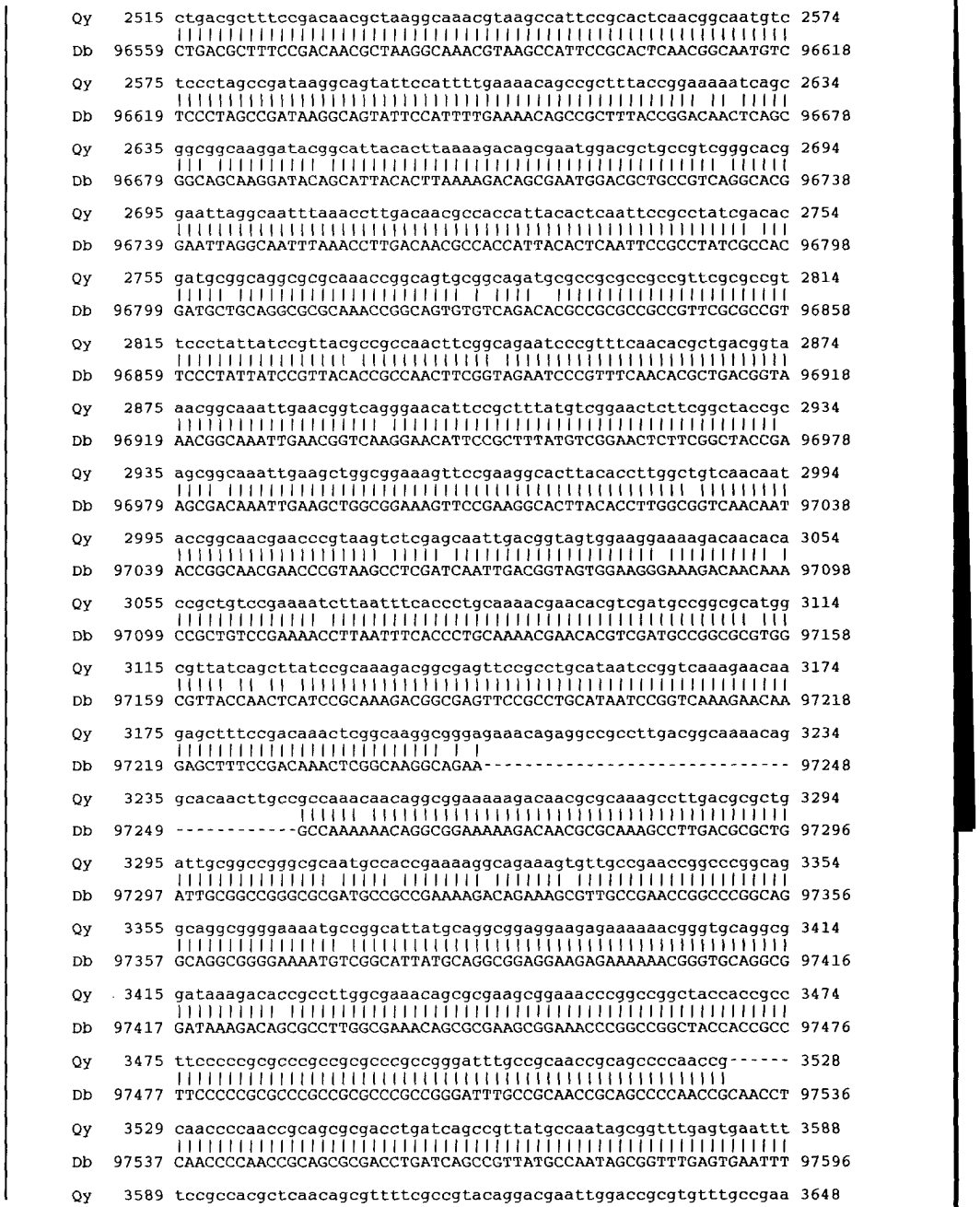
Qy    1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
           || || ||  || | ||  |  | | ||||||||||||||||||||||||||||||||||
Db   95239 GGCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 95298

Qy    1261 gagggtaattttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
           || ||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   95299 GAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 95358

Qy    1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
           |||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||
Db   95359 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 95418

Qy    1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
           |||||||||||||||||||  ||||||||||||||||||||||||||| |||||
Db   95419 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 95478
```





```
Db  97597 TCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAA 97656

Qy   3649 gaccgccgcaacgccgtttggacaagcggcatccgggacaccaaaacactaccgttcgcaa 3708
Db  97657 GACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAA 97716

Qy   3709 gatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctc 3768
Db  97717 GATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC 97776

Qy   3769 ggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgac 3828
Db  97777 GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGAC 97836

Qy   3829 ggcatcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggc 3888
Db  97837 GGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGC 97896

Qy   3889 aggttcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggc 3948
Db  97897 AGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCAGTCTTTCAGACGGC 97956

Qy   3949 atcagaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgca 4008
Db  97957 ATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCC 98016

Qy   4009 ggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaa 4068
Db  98017 GGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAA 98076

Qy   4069 gcggattaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctac 4128
Db  98077 GCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGCTAC 98136

Qy   4129 cgcgcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgcct 4188
Db  98137 CGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCT 98196

Qy   4189 tatttgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaatacc 4248
Db  98197 TATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACC 98256

Qy   4249 gccgtattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaa 4308
Db  98257 GCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAA 98316

Qy   4309 atcaaaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcg 4368
Db  98317 ATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCG 98376

Qy   4369 cagcacagcgcgggcatcaaattaggctaccgctggtaa 4407
Db  98377 CAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 98415


RESULT  10
A96618
LOCUS       A96618                 4350 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 651 from Patent WO9924578.
ACCESSION   A96618
VERSION     A96618.1  GI:6780201
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4350)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 651 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4350
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     1225 a    1161 c    1060 g     842 t      62 others
ORIGIN


  Query Match             76.8%;  Score 3385.8;  DB 6;  Length 4350;
  Best Local Similarity   86.3%;  Pred. No. 0;
  Matches 3814;  Conservative    0;  Mismatches  524;  Indels   81;  Gaps    7;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
Db     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aatgttgattttggtgcggagggaagcaatcccgatcagcaccgctttttcttaccaaatt 420
Db     361 AACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAAATT 420

Qy     421 gtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcat 480
Db     421 GTGAAAAGAAATAATTATAAGCCTG---ACAATTCACACCCTTACAACGGCGATTANCAT 477

Qy     481 atgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatg 540
Db     478 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATG 537

Qy     541 gatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggc 600
Db     538 AGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGA 597

Qy     601 agacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatcatatt 660
Db     598 CACCACTATTGGCGTTATGATGATGACAAACACGGCG------------------ATTTA 639

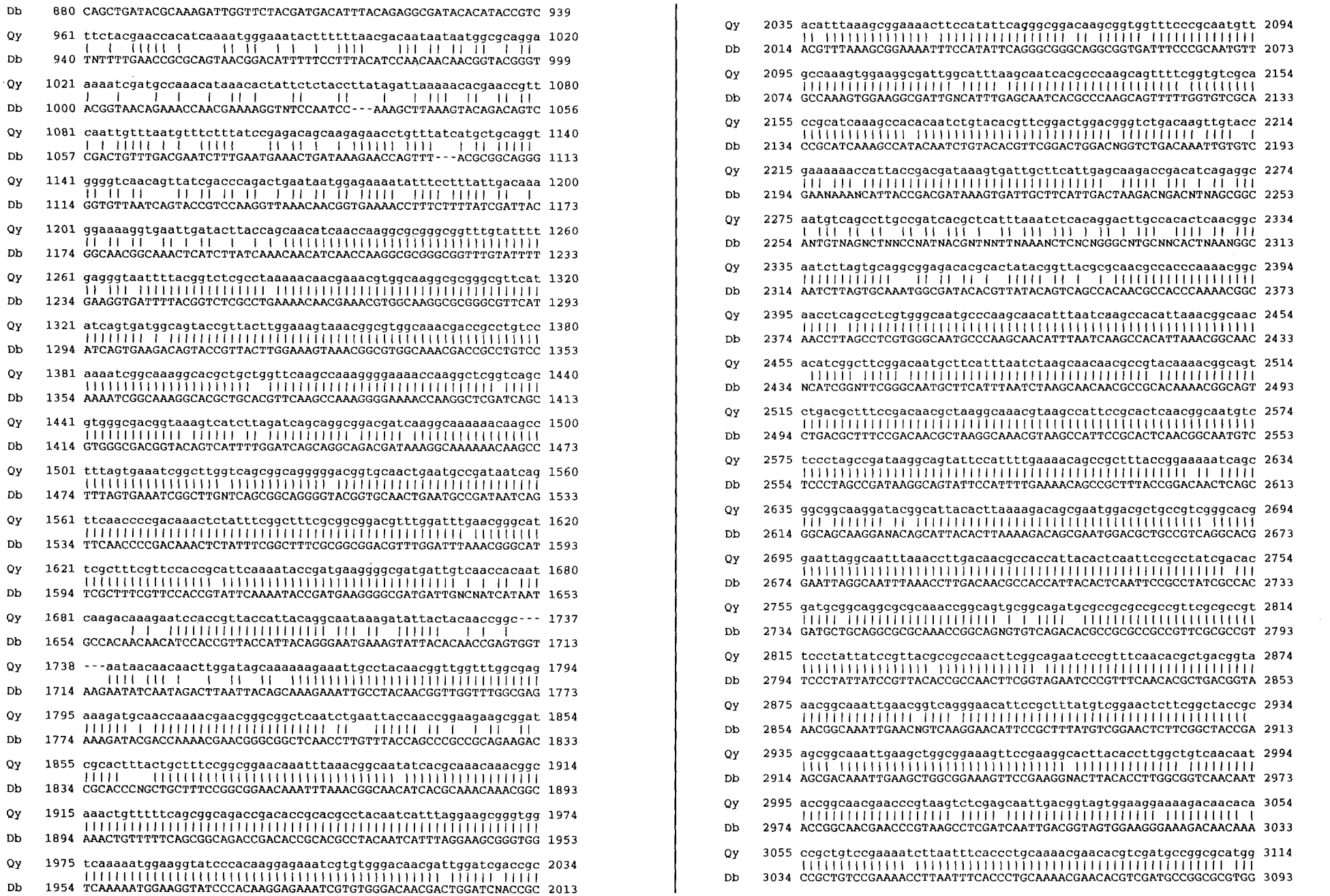
Qy     661 gcaagcgcatattcttggctcgtcggtggcaataccttgcacaaaatggatcaggtggt 720
Db     640 TCCTACTCCGGCGCATGGTTAATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAAT 699

Qy     721 ggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttaccaaca 780
Db     700 GGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATT 759

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
Db     760 GCAGGTGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAA 819

Qy     841 tggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatggcttc 900
Db     820 TGGCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTC 879

Qy     901 cagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
```


Db 880 CAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACATACCGTC 939
Qy 961 ttctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgcagga 1020
Db 940 TNTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACAACGGTACGGGT 999
Qy 1021 aaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaaccgtt 1080
Db 1000 ACGGTAACAGAAACCAACGAAAAGGTNTCCAATCC---AAAGCTTAAAGTACAGACAGTC 1056
Qy 1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db 1057 CGACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTT---ACGCGGCAGGG 1113
Qy 1141 ggggtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacaaa 1200
Db 1114 GGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGATTAC 1173
Qy 1201 ggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatttt 1260
Db 1174 GGCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 1233
Qy 1261 gagggtaatttttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgttcat 1320
Db 1234 GAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1293
Qy 1321 atcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
Db 1294 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1353
Qy 1381 aaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagc 1440
Db 1354 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1413
Qy 1441 gtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcc 1500
Db 1414 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1473
Qy 1501 tttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcag 1560
Db 1474 TTTAGTGAAATCGGCTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1533
Qy 1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcat 1620
Db 1534 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1593
Qy 1621 tcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
Db 1594 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGNCNATCATAAT 1653
Qy 1681 caagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccggc--- 1737
Db 1654 GCCACAACAACATCCACCGTTACCATTACAGGGAATGAAAGTATTACACAACCGAGTGGT 1713
Qy 1738 ---aataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
Db 1714 AAGAATATCAATAGACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1773
Qy 1795 aaagatgcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggat 1854
Db 1774 AAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGAC 1833
Qy 1855 cgcactttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggc 1914
Db 1834 CGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGC 1893
Qy 1915 aaactgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtgg 1974
Db 1894 AAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGG 1953
Qy 1975 tcaaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgc 2034
Db 1954 TCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNACCGC 2013
Qy 2035 acatttaaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgtt 2094
Db 2014 ACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTT 2073
Qy 2095 gccaaagtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgca 2154
Db 2074 GCCAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCA 2133
Qy 2155 ccgcatcaaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtacc 2214
Db 2134 CCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACNGGTCTGACAAATTGTGTC 2193
Qy 2215 gaaaaaaccattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggc 2274
Db 2194 GAANAAANCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACNGACNTNAGCGGC 2253
Qy 2275 aatgtcagccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggc 2334
Db 2254 ANTGTNAGNCTNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGC 2313
Qy 2335 aatcttagtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggc 2394
Db 2314 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGC 2373
Qy 2395 aacctcagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2454
Db 2374 AACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 2433
Qy 2455 acatcggcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagt 2514
Db 2434 NCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACAAAACGGCAGT 2493
Qy 2515 ctgacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtc 2574
Db 2494 CTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTC 2553
Qy 2575 tccctagccgataaggcagtattccattttgaaaacagccgctttaccggaaaaatcagc 2634
Db 2554 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCAGC 2613
Qy 2635 ggcggcaaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacg 2694
Db 2614 GGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACG 2673
Qy 2695 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacac 2754
Db 2674 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2733
Qy 2755 gatgcggcaggcgcgcaaaccggcagtgcggcagatgcgccgcgccgccgttcgcgccgt 2814
Db 2734 GATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGT 2793
Qy 2815 tccctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacggta 2874
Db 2794 TCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTA 2853
Qy 2875 aacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctaccgc 2934
Db 2854 AACGGCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGA 2913
Qy 2935 agcggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaacaat 2994
Db 2914 AGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTTGGCGGTCAACAAT 2973
Qy 2995 accggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaacaca 3054
Db 2974 ACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAAGACAACAAA 3033
Qy 3055 ccgctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgcatgg 3114
Db 3034 CCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGG 3093

```
Qy   3115 cgttatcagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaa 3174
          ||||| || || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3094 CGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA 3153

Qy   3175 gagctttccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaacag 3234
          |||||||||||||||||||||||||| | |
Db   3154 GAGCTTTCCGACAAACTCGGCAAGGCAGAA------------------------------ 3183

Qy   3235 gcacaacttgccgccaaacaacaggcggaaaaagacaacgcgcaaagccttgacgcgctg 3294
                    |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3184 ------------GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTG 3231

Qy   3295 attgcggccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccggcag 3354
          |||||||||||||||| ||||| |||||||| ||||||| |||||||||||||||||| |
Db   3232 ATTGCGGCCGGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCNG 3291

Qy   3355 gcaggcggggaaaatgccggcattatgcaggcggaggaagagaaaaaacgggtgcaggcg 3414
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3292 GCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCG 3351

Qy   3415 gataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccaccgcc 3474
          |||||||||| ||| ||||||||||||||||||||||||||||||||||| ||||||||||
Db   3352 GATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACCACCGCC 3411

Qy   3475 ttccccccgcgcccgccgcgcccgccgggatttgccgcaaccgcagccccaaccg------ 3528
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3412 TTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCT 3471

Qy   3529 caaccccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaattt 3588
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   3472 CAACCCCAACCGCAGCGCGACCTGATNAGCCGTTATGCCAATAGCGGTTTGAGTGAATTT 3531

Qy   3589 tccgccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaa 3648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3532 TCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAA 3591

Qy   3649 gaccgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaa 3708
          |||||||||||||| |||||||||||| |||||||| ||||||||||||||||||||||
Db   3592 GACCGCCGCAACGCNGTTTGGACAAGCNGCATCCGGNACACCAAACACTACCGTTCGCAA 3651

Qy   3709 gatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctc 3768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3652 GATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC 3711

Qy   3769 ggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgac 3828
          |||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||
Db   3712 GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACANCTTCGACGAC 3771

Qy   3829 ggcatcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggc 3888
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   3772 GGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGC 3831

Qy   3889 aggttcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggc 3948
          |||||||||||||||||| |||||||||||||| ||||||  || |||||||||||||||
Db   3832 AGGTTCGACATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCANTCTNTCAGACGGC 3891

Qy   3949 atcagaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgca 4008
          ||| |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   3892 ATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCC 3951

Qy   4009 ggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaa 4068
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   3952 GGTTTCGGCGGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAA 4011

Qy   4069 gcggattaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctac 4128
          |||||||||||| |||||||||||||||||||||||| || ||||| |||||||| |||
Db   4012 GCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGNTAC 4071

Qy   4129 cgcgcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgcct 4188
          || ||||||||||||||||||||||||||||||||||||||||||| |||||||| |||
Db   4072 CGNGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATNTCCATCACNCCT 4131

Qy   4189 tatttgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaatacc 4248
          ||||| |||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db   4132 TATTTNAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACC 4191

Qy   4249 gccgtattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaa 4308
          || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4192 GCNGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAA 4251

Qy   4309 atcaaaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcg 4368
          |||||||||||||||||||| |||||||||||||||||| || |||||| ||||||||
Db   4252 ATCAAAGGTTTCACGCTGTCCNTCCACGCTGCCGCCGCCAAAGGNCCGCAACTGGAAGCG 4311

Qy   4369 cagcacagcgcgggcatcaaattaggctaccgctggtaa 4407
          || ||||||||||||||||||||||||||||||||||||
Db   4312 CAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4350


RESULT  11
A96614
LOCUS       A96614                  4355 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 647 from Patent WO9924578.
ACCESSION   A96614
VERSION     A96614.1  GI:6780199
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4355)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 647 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4355
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      840 a    775 c    694 g    616 t   1430 others
ORIGIN


  Query Match             56.8%;  Score 2501.6;  DB 6;  Length 4355;
  Best Local Similarity   62.5%;  Pred. No. 0;
  Matches 2780;  Conservative   10;  Mismatches 1524;  Indels  134;  Gaps    7;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagcccctaaaaccggccgc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||| |||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGTCGC 60

Qy      61 atccgcttctcgcccgcttacttagccatatgcctgtcgttcggcattctgccccaagcc 120
           |||||||||||  | ||||||||||||||||||||||||||||||||||||| |||||||||
Db      61 ATCCGCTTCTCNGCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy     121 cgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

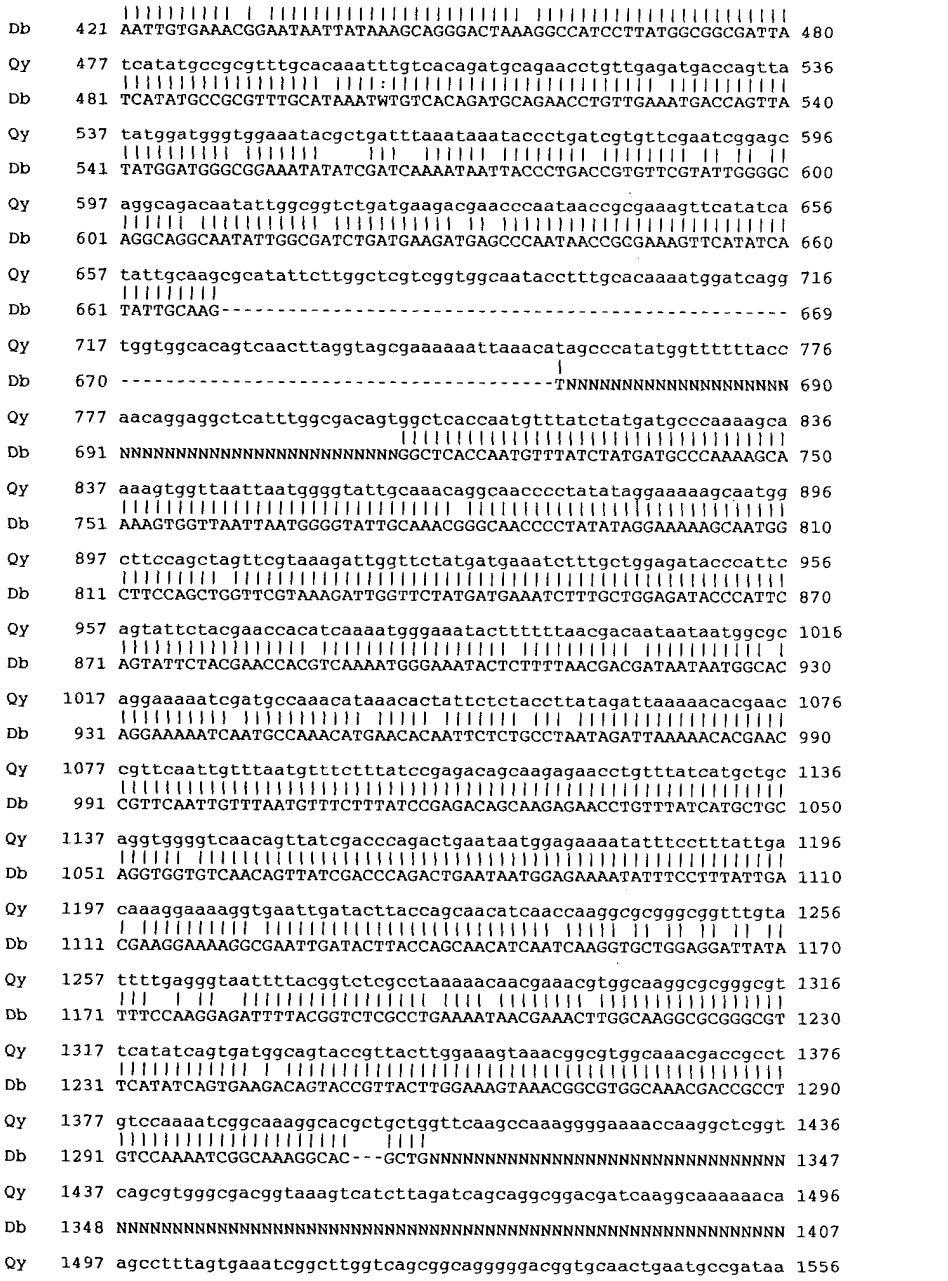
Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggc 300
           |||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggcgggcg-atcaatatattgtgagcgtggcacataacggcggctataa 359
           |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGTATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAA 360

Qy     360 caatgttgattttggtgcggagggaagcaat---cccgatcagcaccgcttttcttacca 416
           ||| |||||||||||||||| |||: |||   || ||||| || || :|| |||| |
Db     361 CAACGTTGATTTTGGTGCGGAAGGAAKNAATATCCCNGATCAACAWCGWWTTACTTATAA 420

Qy     417 aattgtgaaaagaaataattataaagcagggactaacggccatccttatggcggcgatta 476
```


Db 421 AATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTA 480
Qy 477 tcatatgccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagtta 536
Db 481 TCATATGCCGCGTTTGCATAAATWTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTA 540
Qy 537 tatggatgggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagc 596
Db 541 TATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGC 600
Qy 597 aggcagacaatattggcggtctgatgaagacgaacccaataaccgcgaaagttcatatca 656
Db 601 AGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCA 660
Qy 657 tattgcaagcgcatattcttggctcgtcggtggcaatacctttgcacaaaatggatcagg 716
Db 661 TATTGCAAG---------- 669
Qy 717 tggtggcacagtcaacttaggtagcgaaaaaattaaacatagcccatatggtttttacc 776
Db 670 ----------TNNNNNNNNNNNNNNNNNNN 690
Qy 777 aacaggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagca 836
Db 691 NNNNNNNNNNNNNNNNNNNNNNNGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 750
Qy 837 aaagtggttaattaatggggtattgcaaacaggcaacccctatataggaaaaagcaatgg 896
Db 751 AAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGG 810
Qy 897 cttccagctagttcgtaaagattggttctatgatgaaatctttgctggagatacccattc 956
Db 811 CTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTC 870
Qy 957 agtattctacgaaccacatcaaaatgggaaatacttttttaacgacaataataatggcgc 1016
Db 871 AGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCAC 930
Qy 1017 aggaaaaatcgatgccaaacataaacactattctctaccttatagattaaaaacacgaac 1076
Db 931 AGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAAC 990
Qy 1077 cgttcaattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgc 1136
Db 991 CGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGC 1050
Qy 1137 aggtggggtcaacagttatcgacccagactgaataatggagaaaatatttccttttattga 1196
Db 1051 AGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGA 1110
Qy 1197 caaaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgta 1256
Db 1111 CGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATA 1170
Qy 1257 ttttgagggtaatttttacggtctcgcctaaaaacaacgaaacgtggcaaggcgcgggcgt 1316
Db 1171 TTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGT 1230
Qy 1317 tcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcct 1376
Db 1231 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCT 1290
Qy 1377 gtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggt 1436
Db 1291 GTCCAAAATCGGCAAAGGCAC---GCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1347
Qy 1437 cagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaaca 1496
Db 1348 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1407
Qy 1497 agcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataa 1556
Db 1408 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1467
Qy 1557 tcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgg 1616
Db 1468 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1527
Qy 1617 gcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacca 1676
Db 1528 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1587
Qy 1677 caatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccgg 1736
Db 1588 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1647
Qy 1737 caataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaa 1796
Db 1648 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1707
Qy 1797 agatgcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcg 1856
Db 1708 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1767
Qy 1857 cactttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaa 1916
Db 1768 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1827
Qy 1917 actgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtc 1976
Db 1828 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1887
Qy 1977 aaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcac 2036
Db 1888 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1947
Qy 2037 atttaaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgc 2096
Db 1948 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2007
Qy 2097 caaagtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcacc 2156
Db 2008 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2067
Qy 2157 gcatcaaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccga 2216
Db 2068 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2127
Qy 2217 aaaaaccattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaa 2276
Db 2128 NNNNNNNNNNNNNNNGATAAAGTGACTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2187
Qy 2277 tgtcagccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaa 2336
Db 2188 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAA 2247
Qy 2337 tcttagtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaa 2396
Db 2248 TCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAA 2307
Qy 2397 cctcagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacac 2456
Db 2308 CCKTAGCCTCGTGGNSAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACAC 2367
Qy 2457 atcggcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtct 2516
Db 2368 ATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCT 2427
Qy 2517 gacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctc 2576
Db 2428 GACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTC 2487
Qy 2577 cctagccgataaggcagtattccattttgaaaacagccgctttaccggaaaaatcagcgg 2636
Db 2488 CCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGG 2547

```
Qy   2637 cggcaaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacgga 2696
          ||||||||||||||||||||||||||||||||||||||||||||||||||| || :||||
Db   2548 CGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGARCGGA 2607

Qy   2697 attaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacga 2756
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   2608 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGA 2667

Qy   2757 tgcggcaggcgcgcaaaccggcagtgcggcagatgcgccgcgc---------cgccgttc 2807
          ||||||||| |||||||||||||||||| |||||||||||||  .         ||||||||
Db   2668 TGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2727

Qy   2808 gcgccgttccctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgct 2867
          ||||||||||||||||||:|||||| |||||||||||||| ||||||||||||||||||||||
Db   2728 GCGCCGTTCCCTATTATMCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCT 2787

Qy   2868 gacggtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcgg 2927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2788 GACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGG 2847

Qy   2928 ctaccgcagcggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgt 2987
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||| ||
Db   2848 CTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGT 2907

Qy   2988 caacaataccggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaaga 3047
          ||||||||||||||||||| | ||| ||||| ||||||||||||||||||||||||||||
Db   2908 CAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGA 2967

Qy   3048 caacacaccgctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccgg 3107
          ||||| |||||||||||||||| ||||||||||||||||||||||||||||||||| ||
Db   2968 CAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCAGG 3027

Qy   3108 cgcatgg---------------------------cgttatcagcttatccgcaaagacgg 3140
          ||| |||
Db   3028 CGCGTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3087

Qy   3141 cgagttccgcctgcataatccggtcaaagaacaagagctttccgacaaactcggcaaggc 3200

Db   3088 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3147

Qy   3201 gggagaaacagaggccgccttgacggcaaaacaggcacaacttgccgccaaacaacaggc 3260

Db   3148 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3207

Qy   3261 ggaaaaagacaacgcgcaaagccttgacgcgctgattgcggccgggcgcaatgccaccga 3320

Db   3208 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3267

Qy   3321 aaaggcagaaagtgttgccgaaccggcccggcaggcaggcggggaaaatgccggcattat 3380

Db   3268 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3327

Qy   3381 gcaggcggaggaagagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaaca 3440

Db   3328 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3387

Qy   3441 gcgcgaagcggaaacccggccggctaccaccgccttcccccgcgcccgccgcgcccgccg 3500

Db   3388 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3447

Qy   3501 ggatttgccgcaaccgcagccccaaccgcaaccccaaccgcagcgcgacctgatcagccg 3560

Db   3448 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3507

Qy   3561 ttatgccaatagcggtttgagtgaattttccgccacgctcaacagcgttttcgccgtaca 3620

Db   3508 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3567

Qy   3621 ggacgaattggaccgcgtgtttgccgaagaccgccgcaacgccgtttggacaagcggcat 3680
               || |||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3568 NNNNNNNTTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCAT 3627

Qy   3681 ccgggacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacct 3740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3628 CCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCT 3687

Qy   3741 gcgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgca 3800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3688 GCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCA 3747

Qy   3801 caaccggaccggaaacaccttcgacgacggcatcggcaactcggcacggcttgcccacgg 3860
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   3748 CAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGG 3807

Qy   3861 tgccgttttcgggcaatacggcatcggcaggttcgacatcggcatcag-cgcgggcgcgg 3919
          |||||||||||||||||||||||||| ||||||| |||||||||||| ||||||||||
Db   3808 CGCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGNCGCGGGCGCGG 3867

Qy   3920 gttttagtagcggcagcctttcagacggcatcagaggcaaaatccgccgccgcgtgctgc 3979
          ||||||| |||||||||||||||||||||||| |||::|||:|||||||||||||||||
Db   3868 GTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGSMAAAWTCCGCCGCCGCGTGCTGC 3927

Qy   3980 attacggcattcaggcaagataccgcgcaggtttcggcggattcggcatcgaaccgcaca 4039
          |||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db   3928 ATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACA 3987

Qy   4040 tcggcgcaacgcgctatttcgtccaaaaagcggattaccgatacgaaaacgtcaatatcg 4099
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   3988 TCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCG 4047

Qy   4100 ccaccccgggccttgcattcaaccgctaccgcgcgggcattaaggcagattattcattca 4159
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4048 CCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCA 4107

Qy   4160 aaccggcgcaacacatttccatcacgccttatttgagcctgtcctataccgatgccgctt 4219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4108 AACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTT 4167

Qy   4220 ccggcaaagtccgaacgcgcgtcaataccgccgtattggcgcaggatttcggcaaaaccc 4279
          | ||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db   4168 CGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCC 4227

Qy   4280 gcagtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctg 4339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4228 GCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTG 4287

Qy   4340 ccgccgccaaggggccgcaattggaagcgcagcacagcgcgggcatcaaattaggctacc 4399
          |||||||||| || |||||| |||||||||| ||||||||||||||||||||||||||||
Db   4288 CCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACC 4347

Qy   4400 gctggtaa 4407
          ||||||||
Db   4348 GCTGGTAA 4355
```

```
RESULT  12
U32710/c
LOCUS       U32710                 13818 bp    DNA     linear   BCT 29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 25 of 163 of the complete genome.
ACCESSION   U32710 L42023
VERSION     U32710.1  GI:1573200
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
  ORGANISM  Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 13818)
  AUTHORS   Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
```

```
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
REFERENCE   2  (bases 1 to 13818)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
REFERENCE   3  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The H. influenzae sequence has been updated by R. Fleischmann. New
            database matches have been assigned, product names have been
            improved, and a number of frame shifts have been corrected. We
            gratefully acknowledge the work of Tatusov et. al. We have
            incorporated their annotation into the /notes fields of the
            corresponding H. influenzae genes
REFERENCE   5  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
            Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The whole genome was shifted by 588 nucleotides for a new start
COMMENT     On Sep 30, 1996 this sequence version replaced gi:1220911.
FEATURES             Location/Qualifiers
     source          1. .13818
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            complement(134. .343)
                     /gene="HI0235"
     CDS             complement(134. .343)
                     /gene="HI0235"
                     /note="similar to GB:L29458 SP:P36675 PID:473421 percent
                     identity: 73.81; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21905.1"
                     /db_xref="GI:1573201"
                     /translation="MRKKQKSAVENETVYLHTRGVIKDNAVMALLGDKLFRQRVEKKR
                     KGKGSYQRKAKHPGKIFEKPDYKFF"
     gene            complement(411. .761)
                     /gene="HI0236"
     CDS             complement(411. .761)
                     /gene="HI0236"
                     /note="similar to PID:1805555 percent identity: 45.74;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="arsenate reductase, putative"
                     /protein_id="AAC21906.1"
                     /db_xref="GI:1573202"
                     /translation="MSVIIYHNPHCSKSRETLALLENKGIQPIIELYLQKQYSVNELQ
                     SIAKKLGIDDVRQMMRTKDELYKSLNLDNLDLSQAELFKAISEHSALIERPIVINGDK
                     AKIGRPPETVLEIL"
     gene            840. .1888
                     /gene="HI0237"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     GB:U00096 SP:P77406 PID:1621505 PID:1788838 percent
                     identity: 55.75; identified by sequence similarity;
                     putative"
                     /pseudo
     gene            complement(2003. .2980)
                     /gene="HI0239"
     CDS             complement(2003. .2980)
                     /gene="HI0239"
                     /note="similar to SP:P19674 GB:X56175 PID:581231 GB:U00096
                     PID:1773093 percent identity: 48.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-export membrane protein (secF)"
                     /protein_id="AAC21907.1"
                     /db_xref="GI:1573204"
                     /translation="MMKLFTKDKDGHFIREINGIKLPFPLTEFMKVRKLGYILSALLM
                     VISLFFIITKGFNWGLDFTGGVVFDTHFSQSANLEQIRSKLHENGIESPIVQTTGSVQ
                     DVMIRLPASNNDSTIGEHVKSMLQNVDKDIQIRSIEFVGPNVGEELAQGAVYATLATL
                     AMVLIYVGSRFEWRLGFGSIASLAHDVIITLGVFSALQIEIDLTFVAAILSVVGYSIN
                     DSIVVFDRVRENFRKIRRLDTIDIIDISLTQTLSRTIITSVTTLVVVMALFFFGGPSI
                     HNFSLALLVGIGFGTYSSIFVAIAIAYDVGLRREHMIPPKVDKEIDELP"
     gene            complement(2988. .4838)
                     /gene="HI0240"
     CDS             complement(2988. .4838)
                     /gene="HI0240"
                     /note="similar to SP:P19673 GB:S68715 GB:X56175 PID:581230
                     PID:545175 percent identity: 59.74; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-export membrane protein (secD)"
                     /protein_id="AAC21908.1"
                     /db_xref="GI:1573205"
                     /translation="MLNRYPLWKNLMVIFIVAIGILYSLPNIYGEDPAVQISGTRGQE
                     ANTSVLGQVQDVLKTNNLPTKSIVLENGSILARFTNTDDQLLAKDKIAERLGNNYTTA
                     LNLAPATPAWLSMFGANPMKWGLDLRGGVRFLMEVDMNATLVKRQEQLQDSLRGELRK
                     EKIQYTAIKNTEHFGTLITLANVSQRAKAERIIRQLHPTLDITEPDADSINLGLSTAA
                     LNEARDLAIEQNLTILRKRVAELGVAEAVIQRQGAERIVIELPGVQDTARAKEILGAT
                     ATLEFRIVNQNVTADAISRNMLPADSEVKYDRQGHPVALFKRAVLGGEHIINSSSGLD
                     QHSSTPQVSVTLDSEGGEIMSQTTKKYYKKPMATLYVEYKDNGKKDENGKTILEKHEE
                     VINVATIQGRFGSNFQITGVDSIAEAHNLSTLLKSGALIAPIQIVEERTIGPSLGAQN
                     VEQGINASLWGLVAVIAFMLFYYKMFGVIASFALVINIVLLVGLMSILPGATLSMPGI
                     AGIVLTLGMSVDANVLIFERIKEEIRNGRSIQQAINEGYNGAFTSIFDANLTTILTAI
                     ILYAVGTGPIQGFAITLSLGVAISMFTAITGTRALVNALYGGKQLKKLLI"
     gene            complement(4908. .5201)
                     /gene="HI0241"
     CDS             complement(4908. .5201)
                     /gene="HI0241"
                     /note="similar to PID:599589 percent identity: 62.37;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21909.1"
                     /db_xref="GI:1573206"
                     /translation="MEAQSPMSTLFIFVIFGLIFYFMIYRPQAKRNKEHKKLMSELAK
                     GTEVLTAGGVIGKITKVTEGSDSIVIALNDTTEITINRNYIVSVLPKGSLKSL"
     gene            complement(5309. .5530)
                     /gene="HI0242"
     CDS             complement(5309. .5530)
                     /gene="HI0242"
                     /note="similar to GB:L77117 SP:Q58397 PID:1591652 percent
                     identity: 35.71; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21910.1"
                     /db_xref="GI:1573207"
                     /translation="MKYQLNLTALRCPIPLLSAKKALKNLDKNDELMLILNLESAVEN
                     FSIFAEENSVALVEQYYASEKEFIVILKK"
     gene            complement(5527. .6045)
                     /gene="HI0243"
```

```
     CDS             complement(5527. .6045)
                     /gene="HI0243"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. influenzae predicted coding region HI0243"
                     /protein_id="AAC21913.1"
                     /db_xref="GI:1573208"
                     /translation="MQILEPQQFATWNEPIEMLYACHSKVKRFCRQLSILPDYLEKHG
                     YTQAVLNDVEQILSYFNRAAPLHHDDEELDFFPQLVKVAPQTQTSIDELEKQHEYLHE
                     NWNALSVQLEELISEQRQDIDKHLIERFIQGYDRHIALEEPLFEMGRECLSADILTEM
                     GKHMSARRQVKE"
     gene            complement(6110. .7258)
                     /gene="HI0244"
     CDS             complement(6110. .7258)
                     /gene="HI0244"
                     /note="similar to GB:M63939 SP:P19675 GB:S68715 GB:X56175
                     PID:147966 percent identity: 81.60; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tRNA-guanine transglycosylase (tgt)"
                     /protein_id="AAC21911.1"
                     /db_xref="GI:1573209"
                     /translation="MKYELDKTSGNARRGRLVFERPQGTFSVETPAFMPVGTYGTVKG
                     MTPEEVRATGAEILLGNTFHLWLRPGQEVMRKHGDLHDFMQWHRPILTDSGGFQVFSL
                     GKLRKITEEGVKFQNPINGERIFLSPEKSMEIQYDLGSDIVMIFDECTPYPATFDYAK
                     KSMEMSLRWAKRSRDRFDELGNKNALFGIIQGGVFEELRKVSLEGLVNIGFDGYAVGG
                     LAVGEPKEDMHRILEYICPQIPADKPRYLMGVGKPEDLVEGVRRGIDMFDCVMPTRNA
                     RNGHLFVTDGIVKIRNAKYRDDTSPLDPECDCYTCKNYTKAYLYHLDKCGEILGARLN
                     TIHNLRYYQRLMAEIRQAIEDDRFDDFVVEFYARMGKPVPPLQLADKS"
     gene            complement(7602. .8693)
                     /gene="HI0245"
     CDS             complement(7602. .8693)
                     /gene="HI0245"

  Query Match             30.1%;  Score 1327;  DB 1;  Length 13818;
  Best Local Similarity   59.0%;  Pred. No. 1.8e-300;
  Matches 2569;  Conservative    0;  Mismatches 1610;  Indels  176;  Gaps   10;

Qy      78 ttacttagccatatgcctgtcgttcggcattctgccccaagcccgggcgggacacactta 137
           ||  ||| ||   ||  | || || || ||     | |||||| |||| || || |||||
Db   13605 TTTTTTAACCGCTTGTGTTTCATTAGGGATAGCATCACAAGCCTGGGCAGGTCATACTTA 13546

Qy     138 tttcggcatcaactaccaatactatcgcgactttgccgaaaataaaggcaagtttgcagt 197
           ||| || ||  |||||||||| ||||| || |||||||| |||||||| |||||  ||||
Db   13545 TTTTGGGATTGACTACCAATATTATCGTGATTTTGCCGAGAATAAAGGGAAGTTCACAGT 13486

Qy     198 cggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatcgatgac 257
           ||||| ||| |||||||||||||| |||||| ||||| | || || |||| ||| |||||
Db   13485 TGGGGCTAAAAATATTGAGGTTTATAACAAAGAAGGGCAATTAGTTGGCACATCAATGAC 13426

Qy     258 gaaagccccgatgattgatttttctgtggtatcgcgtaacggcgtggcggcattggcggg 317
           |||||||||||||||||||||||| ||||| |||||||||||||||||||||||| |  ||
Db   13425 AAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCATTAGTAGG 13366

Qy     318 cgatcaatatattgtgagcgtggcacataacggcggctataacaatgttgattttggtgc 377
           |||||| |||||||||||||||||||||||||||||| |||||| ||||||||||||||
Db   13365 CGATCAGTATATTGTGAGCGTGGCACATAACGGCGGATATAACGATGTTGATTTTGGTGC 13306

Qy     378 ggagggaagcaatcccgatcagcaccgcttttcttaccaaattgtgaaaagaaataatta 437
           || ||| | || || |||||||||||||||| |||| |||||||| ||||||||||||||
Db   13305 AGAAGGACGAAACCCTGATCAGCACCGCTTTACTTATCAAATTGTAAAAAGAAATAATTA 13246

Qy     438 taaagcagggactaacggccatccttatggcggcgattatcatatgccgcgtttgcacaa 497
           | ||||  ||   |     ||||||||||  || ||||||||||||||| ||||| || ||
Db   13245 TCAAGCTTGGGAGAGAAAGCATCCTTATGATGGAGATTATCATATGCCTCGTTTACATAA 13186

Qy     498 atttgtcacagatgcagaacctgttgagatgaccagttatatggatgggtggaaatacgc 557
           |||||| || || || |||||||| |  |||||  |   ||||||||||      ||| ||
Db   13185 ATTTGTAACTGAAGCTGAACCTGTGGGTATGACAACAAATATGGATGGAAAAGTATATGC 13126

Qy     558 tgatttaaataaataccctgatcgtgttcgaatcggagcaggcagacaatattggcggtc 617
           ||||  | | || || ||||| ||||| || || ||  ||||  | || |||||||||  |
Db   13125 TGATAGAGAGAACTATCCTGAGCGTGTACGTATAGGCTCAGGACGTCAGTATTGGCGTAC 13066

Qy     618 tgatgaagacgaacccaataaccgcgaaagttcatatcatattgcaagcgcatattcttg 677
           ||| |||| |||   |  ||     | |||||||||| || |  || | ||||||  ||
Db   13065 AGATAAAGATGAAGAAACGAATGTACATAGTTCATATTATGTCTCAGGTGCATATCGTTA 13006

Qy     678 gctcgtcggtggcaataccttgcacaaaatggatcaggtggtggcacagtcaacttagg 737
           ||    |  || |||||  | | || | ||||   |||  ||| ||||||||  |  |
Db   13005 TCTTACTGCAGGAAATACCCATACTCAGAGTGGAAATGGTAATGGTACAGTCAATCTTAG 12946

Qy     738 tagcgaaaaaattaaacatagcccatatggttttttaccaacaggaggctcatttggcga 797
           | |  |   | |||  | ||  |  ||||||   |||||||| || || ||    |||||
Db   12945 TGGTAATGTAGTTAGCCCTAATCATTATGGTCCATTACCAACGGGTGGTTCTAAAGGCGA 12886

Qy     798 cagtggctcaccaatgtttatctatgatgcccaaaagcaaaagtggttaattaatggggt 857
           || || || |||||||||||| ||||||||  | ||| || | ||| | || ||||  ||
Db   12885 TAGCGGTTCGCCAATGTTTATTTATGATGCGAAGAAGAAACAATGGCTTATAAATGCTGT 12826

Qy     858 attgcaaacaggcaacccctatataggaaaaagcaatggcttccagctagttcgtaaaga 917
           ||| ||||| ||  | || | | | |||| | | ||||| || ||| || | ||| ||||
Db   12825 ATTACAAACTGGGCATCCTTTTTTCGGAAGAGGTAATGGGTTTCAGTTAATACGTGAAGA 12766

Qy     918 ttggttctatgatgaaatctttgctggagatacccattcagtatt------ctacgaacc 971
           |||||| ||| ||||| |  |||| |  ||||||| |   || ||       |||   ||
Db   12765 ATGGTTTTATAATGAAGTTCTTGCGGTTGATACCCCTAGTGTTTTTCAACGCTATATTCC 12706

Qy     972 acatcaaaatgggaaatacttttttaacgacaataataatggcgcaggaaaaatcgat-- 1029
           |    |||||| | || |  |||      |||||| |||| |||| ||| |   |
Db   12705 CCCAATAAATGGACATTATTCCTTTGTATCAAATAATGATGGTACAGGTAAATTAACTTT 12646

Qy    1030 -gccaaacataaacactattctctaccttatagattaaaaacacgaaccgttcaattgtt 1088
            | | || ||   |  ||          | || | ||  | |||| ||  | || ||
Db   12645 AACTAGACCTAGTAAAGATGGCTCTAAAGCAAAATCAGAAGTAGGAACTGTGAAGTTATT 12586

Qy    1089 taatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggtggggtcaa 1148
           ||||   || |||    | || || | ||||| |||| |   || ||||  ||    ||
Db   12585 TAATCCATCGTTAAATCAAACCGCTAAAGAACATGTTAAA---GCAGCAGCAGGCTATAA 12529

Qy    1149 cagttatcgacccagactgaataatggagaaaatatttccttttattgacaaaggaaaagg 1208
           | ||| | ||| ||| || |  |||||  |||||||||  | ||   ||| ||||||||||
Db   12528 TATTTACCAACCAAGAATGGAATATGGAAAAAATATTTACCTTGGCGACCAAGGAAAAGG 12469

Qy    1209 tgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtattttgagggtaa 1268
             || | |    |  || || || ||||| || || || || || || ||||| |||||
Db   12468 AACTTTAACAATCGAAAATAATATAAATCAAGGTGCTGGTGGATTATACTTTGAAGGTAA 12409

Qy    1269 ttttacggtctcgccta---aaaacaacgaaacgtggcaaggcgcgggcgttcatatcag 1325
           ||||   ||       |   |||| ||   ||| |||||||| || |||||   |||  |
Db   12408 TTTTGTTGTAAAAGGCAAGCAAAATAATATAACTTGGCAAGGTGCAGGCGTATCTATTGG 12349

Qy    1326 tgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaat 1385
           | |     || |||  |||||||  ||       | ||| || ||  | || |||||
Db   12348 ACAAGATGCAACTGTTGAATGGAAAGTTCACAATCCTGAAAATGATCGTTTATCTAAAAT 12289

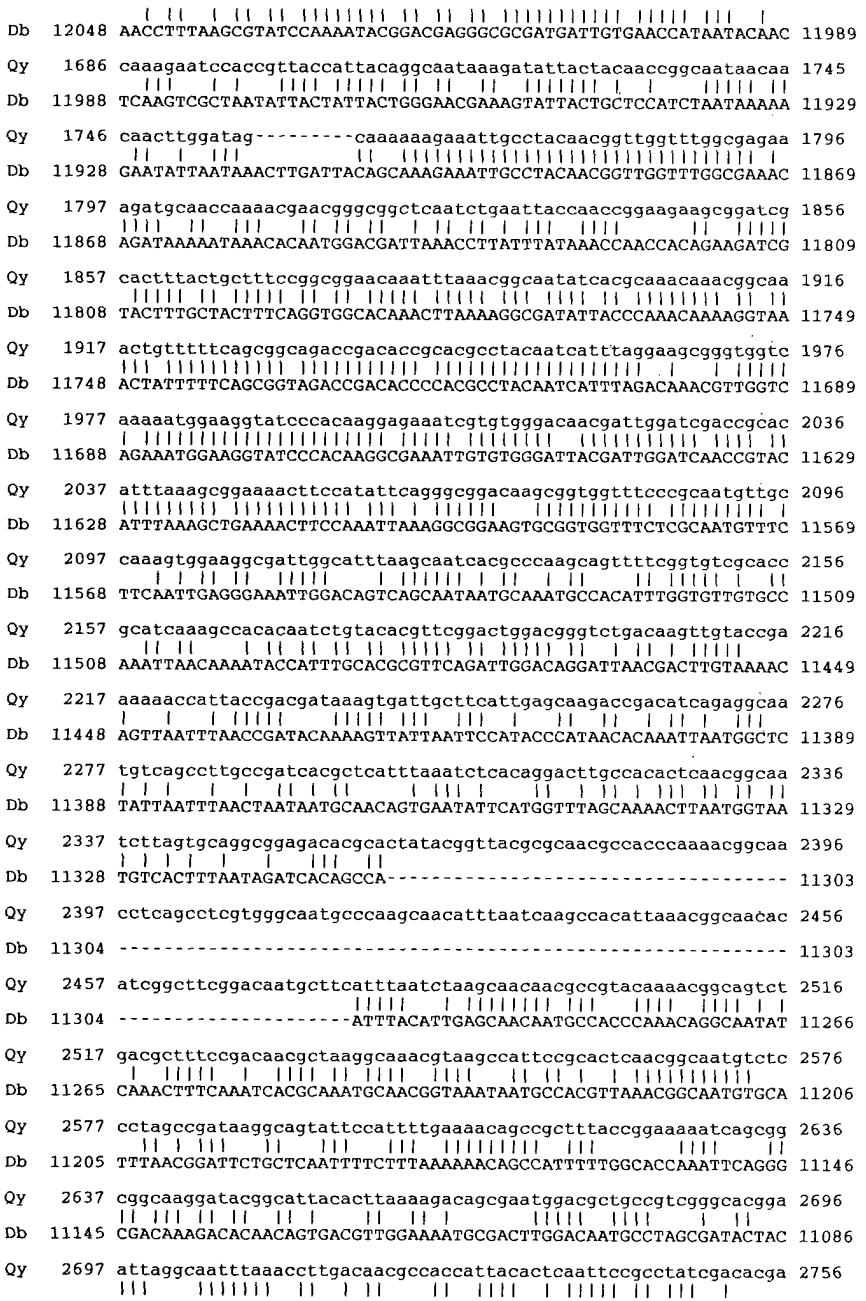
Qy    1386 cggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcggtcagcgtggg 1445
           || | ||||||  |  | ||  | |  || ||  | ||   |||     | || | |||
Db   12288 TGGTATAGGCACTTTATTAGTCAATGGTAAGGGAAAGAATTTAGGAAGTTTAAGTGCGGG 12229

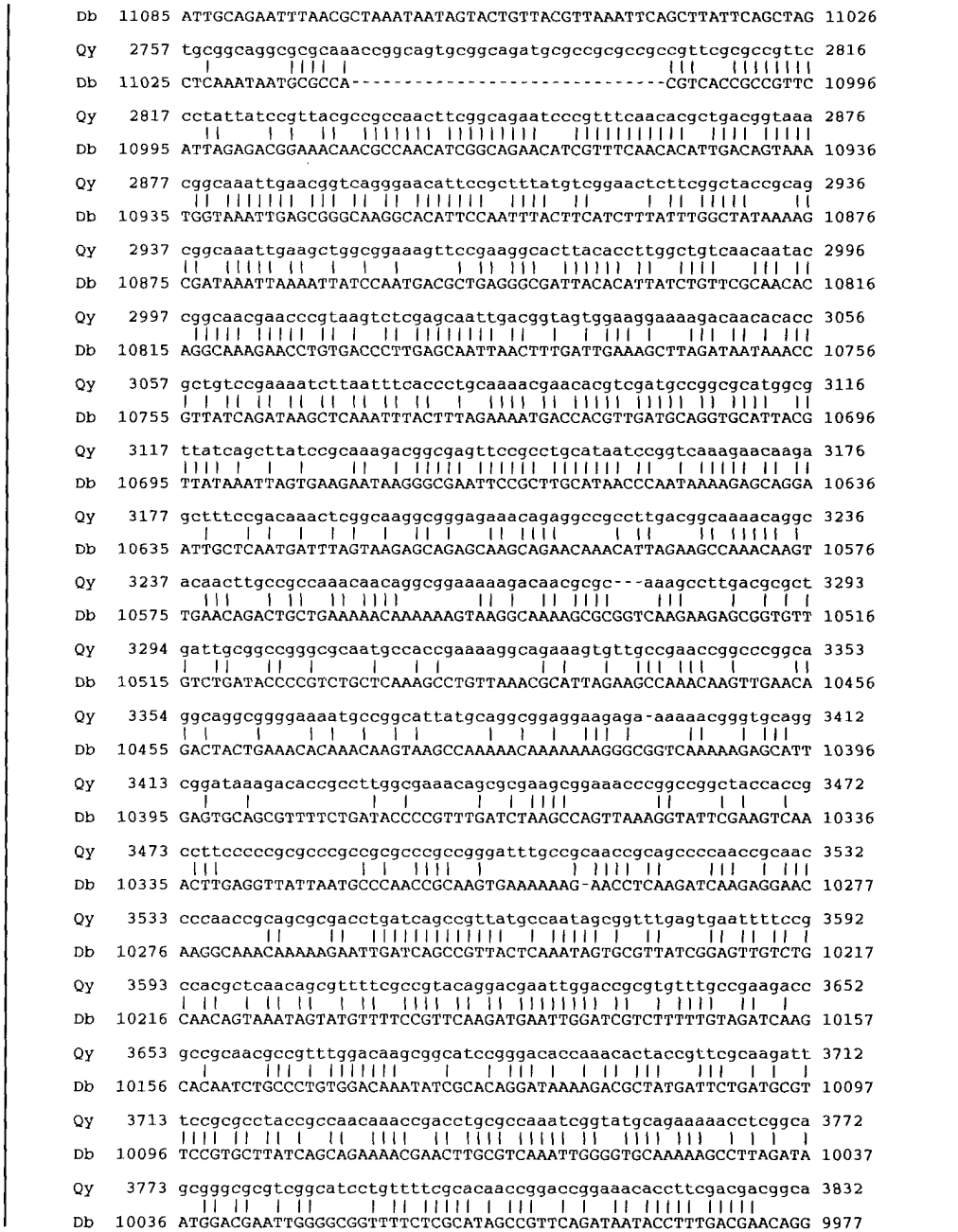
Qy    1446 cgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaacaagcctttag 1505
           |||| |||||||  |||||| || || || ||    ||  ||||||||| || |
Db   12228 TAACGGCAAAGTCATTCTAGATCAACAAGCAGATGAAGCGGGTCAAAAACAAGCTTTCAA 12169

Qy    1506 tgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgataatcagttcaa 1565
           ||| | ||| | || |||||  | |  || || ||| | |||       ||||  |  |
Db   12168 AGAAGTTGGCATTGTAAGCGGTCGAGCAACAGTTCAATTAAATAGTACAGATCAAGTTGA 12109

Qy    1566 ccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacgggcattcgct 1625
           ||  ||||  |||||||||| ||||| || || || || ||| | ||||||||||  |
Db   12108 TCCTAACAATATCTATTTCGGATTTCGTGGTGGTCGCTTAGATCTTAACGGGCATTCATT 12049

Qy    1626 ttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaaccacaatcaaga 1685
```


Db 12048 AACCTTTAAGCGTATCCAAAATACGGACGAGGGCGCGATGATTGTGAACCATAATACAAC 11989
Qy 1686 caaagaatccaccgttaccattacaggcaataaagatattactacaaccggcaataacaa 1745
Db 11988 TCAAGTCGCTAATATTACTATTACTGGGAACGAAAGTATTACTGCTCCATCTAATAAAAA 11929
Qy 1746 caacttggatag---------caaaaaagaaattgcctacaacggttggtttggcgagaa 1796
Db 11928 GAATATTAATAAACTTGATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAAAC 11869
Qy 1797 agatgcaaccaaaacgaacgggcggctcaatctgaattaccaaccggaagaagcggatcg 1856
Db 11868 AGATAAAAATAAACACAATGGACGATTAAACCTTATTTATAAACCAACCACAGAAGATCG 11809
Qy 1857 cactttactgctttccggcggaacaaatttaaacggcaatatcacgcaaacaaacggcaa 1916
Db 11808 TACTTTGCTACTTTCAGGTGGCACAAACTTAAAAGGCGATATTACCCAAACAAAAGGTAA 11749
Qy 1917 actgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtc 1976
Db 11748 ACTATTTTTCAGCGGTAGACCGACACCCCACGCCTACAATCATTTAGACAAACGTTGGTC 11689
Qy 1977 aaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgattggatcgaccgcac 2036
Db 11688 AGAAATGGAAGGTATCCCACAAGGCGAAATTGTGTGGGATTACGATTGGATCAACCGTAC 11629
Qy 2037 atttaaagcggaaaacttccatattcagggcggacaagcggtggtttcccgcaatgttgc 2096
Db 11628 ATTTAAAGCTGAAAACTTCCAAATTAAAGGCGGAAGTGCGGTGGTTTCTCGCAATGTTTC 11569
Qy 2097 caaagtggaaggcgattggcatttaagcaatcacgcccaagcagttttcggtgtcgcacc 2156
Db 11568 TTCAATTGAGGGAAATTGGACAGTCAGCAATAATGCAAATGCCACATTTGGTGTTGTGCC 11509
Qy 2157 gcatcaaagccacacaatctgtacacgttcggactggacgggtctgacaagttgtaccga 2216
Db 11508 AAATTAACAAAATACCATTTGCACGCGTTCAGATTGGACAGGATTAACGACTTGTAAAAC 11449
Qy 2217 aaaaaccattaccgacgataaagtgattgcttcattgagcaagaccgacatcagaggcaa 2276
Db 11448 AGTTAATTTAACCGATACAAAAGTTATTAATTCCATACCCATAACACAAATTAATGGCTC 11389
Qy 2277 tgtcagccttgccgatcacgctcatttaaatctcacaggacttgccacactcaacggcaa 2336
Db 11388 TATTAATTTAACTAATAATGCAACAGTGAATATTCATGGTTTAGCAAAACTTAATGGTAA 11329
Qy 2337 tcttagtgcaggcggagacacgcactatacggttacgcgcaacgccacccaaaacggcaa 2396
Db 11328 TGTCACTTTAATAGATCACAGCCA---------- 11303
Qy 2397 cctcagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacac 2456
Db 11304 ---------- 11303
Qy 2457 atcggcttcggacaatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtct 2516
Db 11304 ----------ATTTACATTGAGCAACAATGCCACCCAAACAGGCAATAT 11266
Qy 2517 gacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctc 2576
Db 11265 CAAACTTTCAAATCACGCAAATGCAACGGTAAATAATGCCACGTTAAACGGCAATGTGCA 11206
Qy 2577 cctagccgataaggcagtattccattttgaaaacagccgctttaccggaaaaatcagcgg 2636
Db 11205 TTTAACGGATTCTGCTCAATTTTCTTTAAAAAACAGCCATTTTTGGCACCAAATTCAGGG 11146
Qy 2637 cggcaaggatacggcattacacttaaaagacagcgaatggacgctgccgtcgggcacgga 2696
Db 11145 CGACAAAGACACAACAGTGACGTTGGAAAATGCGACTTGGACAATGCCTAGCGATACTAC 11086
Qy 2697 attaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgacacga 2756


Db 11085 ATTGCAGAATTTAACGCTAAATAATAGTACTGTTACGTTAAATTCAGCTTATTCAGCTAG 11026
Qy 2757 tgcggcaggcgcgcaaaccggcagtgcggcagatgcgccgcgccgccgttcgcgccgttc 2816
Db 11025 CTCAAATAATGCGCCA----------CGTCACCGCCGTTC 10996
Qy 2817 cctattatccgttacgccgccaacttcggcagaatcccgtttcaacacgctgacggtaaa 2876
Db 10995 ATTAGAGACGGAAACAACGCCAACATCGGCAGAACATCGTTTCAACACATTGACAGTAAA 10936
Qy 2877 cggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctaccgcag 2936
Db 10935 TGGTAAATTGAGCGGGCAAGGCACATTCCAATTTACTTCATCTTTATTTGGCTATAAAAG 10876
Qy 2937 cggcaaattgaagctggcggaaagttccgaaggcacttacaccttggctgtcaacaatac 2996
Db 10875 CGATAAATTAAAATTATCCAATGACGCTGAGGGCGATTACACATTATCTGTTCGCAACAC 10816
Qy 2997 cggcaacgaacccgtaagtctcgagcaattgacggtagtggaaggaaaagacaacacacc 3056
Db 10815 AGGCAAAGAACCTGTGACCCTTGAGCAATTAACTTTGATTGAAAGCTTAGATAATAAACC 10756
Qy 3057 gctgtccgaaaatcttaatttcaccctgcaaaacgaacacgtcgatgccggcgcatggcg 3116
Db 10755 GTTATCAGATAAGCTCAAATTTACTTTAGAAAATGACCACGTTGATGCAGGTGCATTACG 10696
Qy 3117 ttatcagcttatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaaga 3176
Db 10695 TTATAAATTAGTGAAGAATAAGGGCGAATTCCGCTTGCATAACCCAATAAAAGAGCAGGA 10636
Qy 3177 gctttccgacaaactcggcaaggcgggagaaacagaggccgccttgacggcaaaacaggc 3236
Db 10635 ATTGCTCAATGATTTAGTAAGAGCAGAGCAAGCAGAACAAACATTAGAAGCCAAACAAGT 10576
Qy 3237 acaacttgccgccaaacaacaggcggaaaaagacaacgcgc---aaagccttgacgcgct 3293
Db 10575 TGAACAGACTGCTGAAAAACAAAAAAGTAAGGCAAAAGCGCGGTCAAGAAGAGCGGTGTT 10516
Qy 3294 gattgcggccgggcgcaatgccaccgaaaaggcagaaagtgttgccgaaccggcccggca 3353
Db 10515 GTCTGATACCCCGTCTGCTCAAAGCCTGTTAAACGCATTAGAAGCCAAACAAGTTGAACA 10456
Qy 3354 ggcaggcggggaaaatgccggcattatgcaggcggaggaagaga-aaaaacgggtgcagg 3412
Db 10455 GACTACTGAAACACAAACAAGTAAGCCAAAAACAAAAAAAGGGCGGTCAAAAAGAGCATT 10396
Qy 3413 cggataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccaccg 3472
Db 10395 GAGTGCAGCGTTTTCTGATACCCCGTTTGATCTAAGCCAGTTAAAGGTATTCGAAGTCAA 10336
Qy 3473 ccttcccccgcgcccgccgcgcccgccgggatttgccgcaaccgcagccccaaccgcaac 3532
Db 10335 ACTTGAGGTTATTAATGCCCAACCGCAAGTGAAAAAAG-AACCTCAAGATCAAGAGGAAC 10277
Qy 3533 cccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaatttttccg 3592
Db 10276 AAGGCAAACAAAAAGAATTGATCAGCCGTTACTCAAATAGTGCGTTATCGGAGTTGTCTG 10217
Qy 3593 ccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagacc 3652
Db 10216 CAACAGTAAATAGTATGTTTTCCGTTCAAGATGAATTGGATCGTCTTTTTGTAGATCAAG 10157
Qy 3653 gccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagatt 3712
Db 10156 CACAATCTGCCCTGTGGACAAATATCGCACAGGATAAAAGACGCTATGATTCTGATGCGT 10097
Qy 3713 tccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggca 3772
Db 10096 TCCGTGCTTATCAGCAGAAAACGAACTTGCGTCAAATTGGGGTGCAAAAAGCCTTAGATA 10037
Qy 3773 gcgggcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggca 3832
Db 10036 ATGGACGAATTGGGGCGGTTTTCTCGCATAGCCGTTCAGATAATACCTTTGACGAACAGG 9977

```
Qy   3833 tcggcaactcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcaggt 3892
          |    ||    ||     |  |        | | ||| |  |||||       |||   |
Db   9976 TTAAAAATCACGCGACATTAACGATGATGTCGGGTTTTGCCCAATATCAATGGGGCGATT 9917

Qy   3893 tcgacatcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatca 3952
          | |  | ||  | |  | ||| |||||  |||||   | |    |  | || |     |
Db   9916 TACAATTTGGTGTAAATGTGGGTGCGGGAATTAGTGCGAGTAAAATGGCTGAAGAACAAA 9857

Qy   3953 gaggcaaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggtt 4012
          |  | |||||  |  ||    | | |  |||| ||  |  | ||||| || |     |||
Db   9856 GCCGAAAAATTCATCGAAAAGCGATAAATTATGGAGTGAATGCAAGTTATCAGTTCCGTT 9797

Qy   4013 tcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcgg 4072
          | ||   ||| || ||  | ||  |  | ||  |  |  || |||||  |  ||   |
Db   9796 TAGGGCAATTGGGTATTCAGCCTTATTTGGGTGTTAATCGATATTTTATTGAACGTGAAA 9737

Qy   4073 attaccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgcg 4132
          |||| | ||  ||| | || ||  |    || ||| |||||||||| || |||||    |
Db   9736 ATTATCAATCTGAAGAAGTGAAAGTGCAAACACCGAGCCTTGCATTTAATCGCTATAATG 9677

Qy   4133 cgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgccttatt 4192
          | ||||||   |  |||||| |||| |  ||| |  |  | ||   | | | ||||||||
Db   9676 CTGGCATTCGAGTTGATTATACATTTACCCCGACAAATAATATCAGCGTTAAGCCTTATT 9617

Qy   4193 tgagcctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgccg 4252
          |     |   |||  |||| |   ||  || || | |||    || |||| | |
Db   9616 TCTTTGTCAATTATGTTGATGTTTCAAACGCTAACGTACAAACCACTGTAAATAGCACGA 9557

Qy   4253 tattggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatca 4312
          | |||   ||    || ||        |       |||  |||  |||| || |||||
Db   9556 TGTTGCAACAATCATTTGGGCGTTATTGGCAAAAAGAAGTGGGATTAAAGGCAGAAATTT 9497

Qy   4313 aaggtttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcagc 4372
          |  ||||   || ||        | |     |  | ||  |||| | |     ||||
Db   9496 TACATTTCCAACTTTCTGCTTTTATTTCTAAATCTCAAGGTTCGCAACTCGGCAAACAGC 9437

Qy   4373 acagcgcgggcatcaaattaggctaccgctggtaa 4407
          | |  | |||| | ||||| || || || ||||||
Db   9436 AAAATGTGGGCGTGAAATTGGGGTATCGTTGGTAA 9402
```

```
RESULT  13
HIU11024
LOCUS       HIU11024                4317 bp    DNA     linear   BCT 12-JUL-1996
DEFINITION  Haemophilus influenzae N187 adhesion and penetration protein (hap)
            gene, complete cds.
ACCESSION   U11024
VERSION     U11024.1  GI:1408568
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 4317)
  AUTHORS   St. Geme,J.W. III, de la Morena,M.L. and Falkow,S.
  TITLE     A Haemophilus influenzae IgA protease-like protein promotes
            intimate interaction with human epithelial cells
  JOURNAL   Mol. Microbiol. 14 (2), 217-233 (1994)
  MEDLINE   95131744
REFERENCE   2  (bases 1 to 4317)
  AUTHORS   St. Geme,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1994) Joseph W. St. Geme, Molecular Microbiology,
            Washington University School of Medicine, 660 South Euclid Ave.,
            St. Louis, MO 63110, USA
COMMENT     On Jul 12, 1996 this sequence version replaced gi:562013.
FEATURES             Location/Qualifiers
     source          1. .4317
                     /organism="Haemophilus influenzae"
                     /strain="N187"
                     /db_xref="taxon:727"
     gene            61. .4242
                     /gene="hap"
     CDS             61. .4242
                     /gene="hap"
                     /standard_name="Haemophilus adhesion and penetration"
                     /note="HAP"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion and penetration protein"
                     /protein_id="AAB03707.1"
                     /db_xref="GI:1408569"
                     /translation="MKKTVFRLNFLTACISLGIVSQAWAGHTYFGIDYQYYRDFAENK
                     GKFTVGAQNIKVYNKQGQLVGTSMTKAPMIDFSVVSRNGVAALVENQYIVSVAHNVGY
                     TDVDFGAEGNNPDQHRFTYKIVKRNNYKKDNLHPYEDDYHNPRLHKFVTEAAPIDMTS
                     NMNGSTYSDRTKYPERVRIGSGRQFWRNDQDKGDQVAGAYHYLTAGNTHNQRGAGNGY
                     SYLGGDVRKAGEYGPLPIAGSKGDSGSPMFIYDAEKQKWLINGILREGNPFEGKENGF
                     QLVRKSYFDEIFERDLHTSLYTRAGNGVYTISGNDNGQGSITQKSGIPSEIKITLANM
                     SLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDQKQGSLIFASDINQGAGGLYFEG
                     NFTVSPNSNQTWQGAGIHVSENSTVTWKVNGVEHDRLSKIGKGTLHVQAKGENKGSIS
                     VGDGKVILEQQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKFYFGFRGGRLDLN
                     GHSLTFKRIQNTDEGAMIVNHNTTQAANVTITGNESIVLPNGNNINKLDYRKEIAYNG
                     WFGETDKNKHNGRLNLIYKPTTEDRTLLLSGGTNLKGDITQTKGKLFFSGRPTPHAYN
                     HLNKRWSEMEGIPQGEIVWDHDWINRTFKAENFQIKGGSAVVSRNVSSIEGNWTVSNN
                     ANATFGVVPNQQNTICTRSDWTGLTTCQKVDLTDTKVINSIPKTQINGSINLTDNATA
                     NVKGLAKLNGNVTLTNHSQFTLSNNATQIGNIRLSDNSTATVDNANLNGNVHLTDSAQ
                     FSLKNSHFSHQIQGDKGTTVTLENATWTMPSDTTLQNLTLNNSTITLNSAYSASSNNT
                     PRRRSLETETTPTSAEHRFNTLTVNGKLSGQGTFQFTSSLFGYKSDKLKLSNDAEGDY
                     ILSVRNTGKEPETLEQLTLVESKDNQPLSDKLKFTLENDHVDAGALRYKLVKNDGEFR
                     LHNPIKEQELHNDLVRAEQAERTLEAKQVEPTAKTQTGEPKVRSRRAARAAFPDTLPD
                     QSLLNALEAKQAELTAETQKSKAKTKKVRSKRAVFSDPLLDQSLFALEAALEVIDAPQ
                     QSEKDRLAQEEAEKQRKQKDLISRYSNSALSELSATVNSMLSVQDELDRLFVDQAQSA
                     VWTNIAQDKRRYDSDAFRAYQQKTNLRQIGVQKALANGRIGAVFSHSRSDNTFDEQVK
                     NHATLTMMSGFAQYQWGDLQFGVNVGTGISASKMAEEQSRKIHRKAINYGVNASYQFR
                     LGQLGIQPYFGVNRYFIERENYQSEEVRVKTPSLAFNRYNAGIRVDYTFTPTDNISVK
                     PYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKSQGSQ
                     LGKQQNVGVKLGYRW"
BASE COUNT     1496 a    776 c    890 g   1155 t
ORIGIN
```

```
  Query Match             29.7%;  Score 1309.2;  DB 1;  Length 4317;
  Best Local Similarity   59.1%;  Pred. No. 2.2e-296;
  Matches 2570;  Conservative    0;  Mismatches 1568;  Indels  210;  Gaps   11;

Qy      78 ttacttagccatatgcctgtcgttcggcattctgccccaagcccgggcgggacacactta 137
           ||  ||| ||   ||| | || || || ||  |  | |||||  |||| || ||||||||
Db      87 TTTTTTAACCGCTTGCATTTCATTAGGGATAGTATCGCAAGCGTGGGCTGGTCACACTTA 146

Qy     138 tttcggcatcaactaccaatactatcgcgactttgccgaaaataaaggcaagtttgcagt 197
           ||| || ||  | |||||||| ||||| || |||||||| |||||||| |||||  ||||
Db     147 TTTTGGGATTGATTACCAATATTATCGTGATTTTGCCGAGAATAAAGGGAAGTTCACAGT 206

Qy     198 cggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatcgatgac 257
           |||||  || ||||| ||||||| |||||| ||||| | || || |||| ||| |||||
Db     207 TGGGGCTCAAAATATTAAGGTTTATAACAAACAAGGGCAATTAGTTGGCACATCAATGAC 266

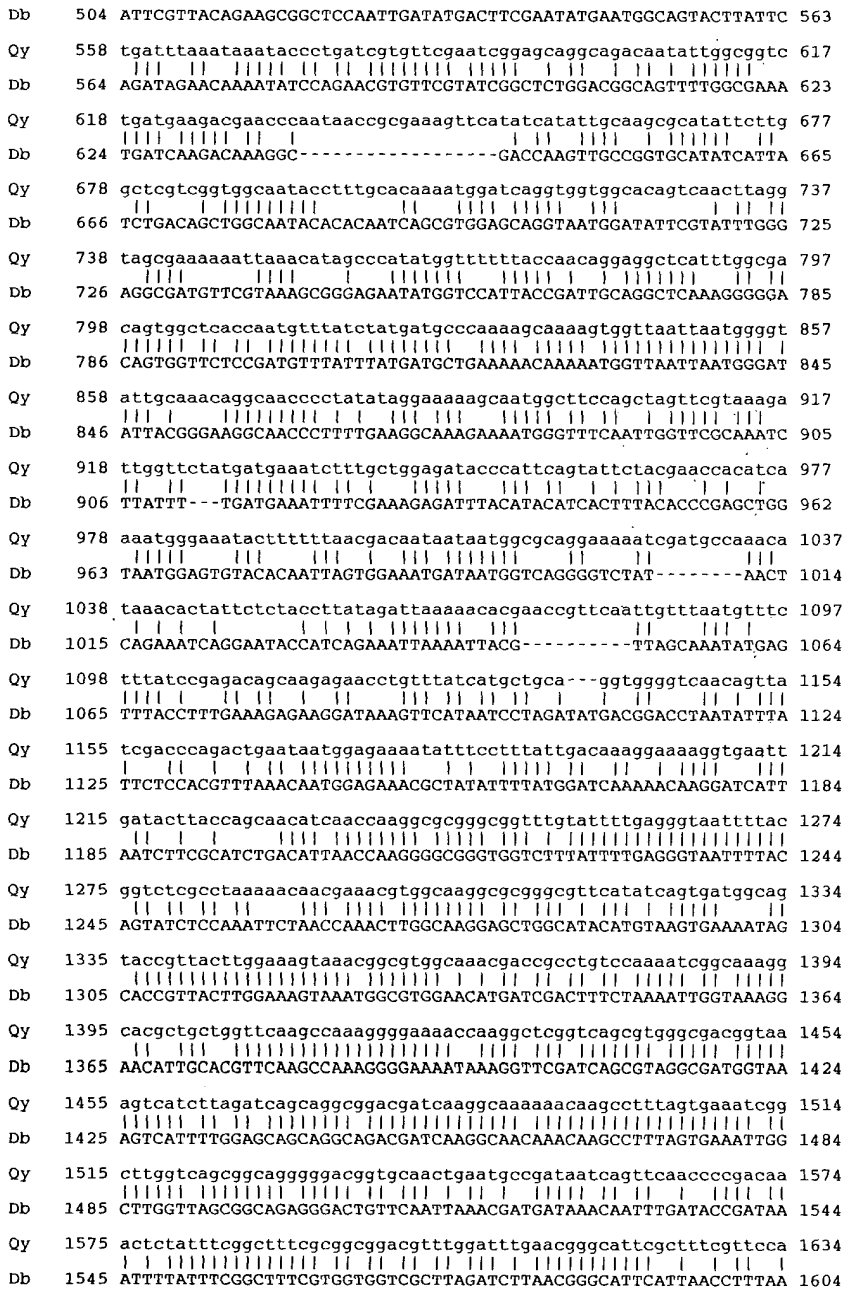
Qy     258 gaaagccccgatgattgatttttctgtggtatcgcgtaacggcgtggcggcattggcggg 317
           |||||||||||||||||||||||||||| || || |||||||||||||| || ||||  |
Db     267 AAAAGCCCCGATGATTGATTTTTCTGTAGTGTCACGTAACGGCGTGGCAGCCTTGGTTGA 326

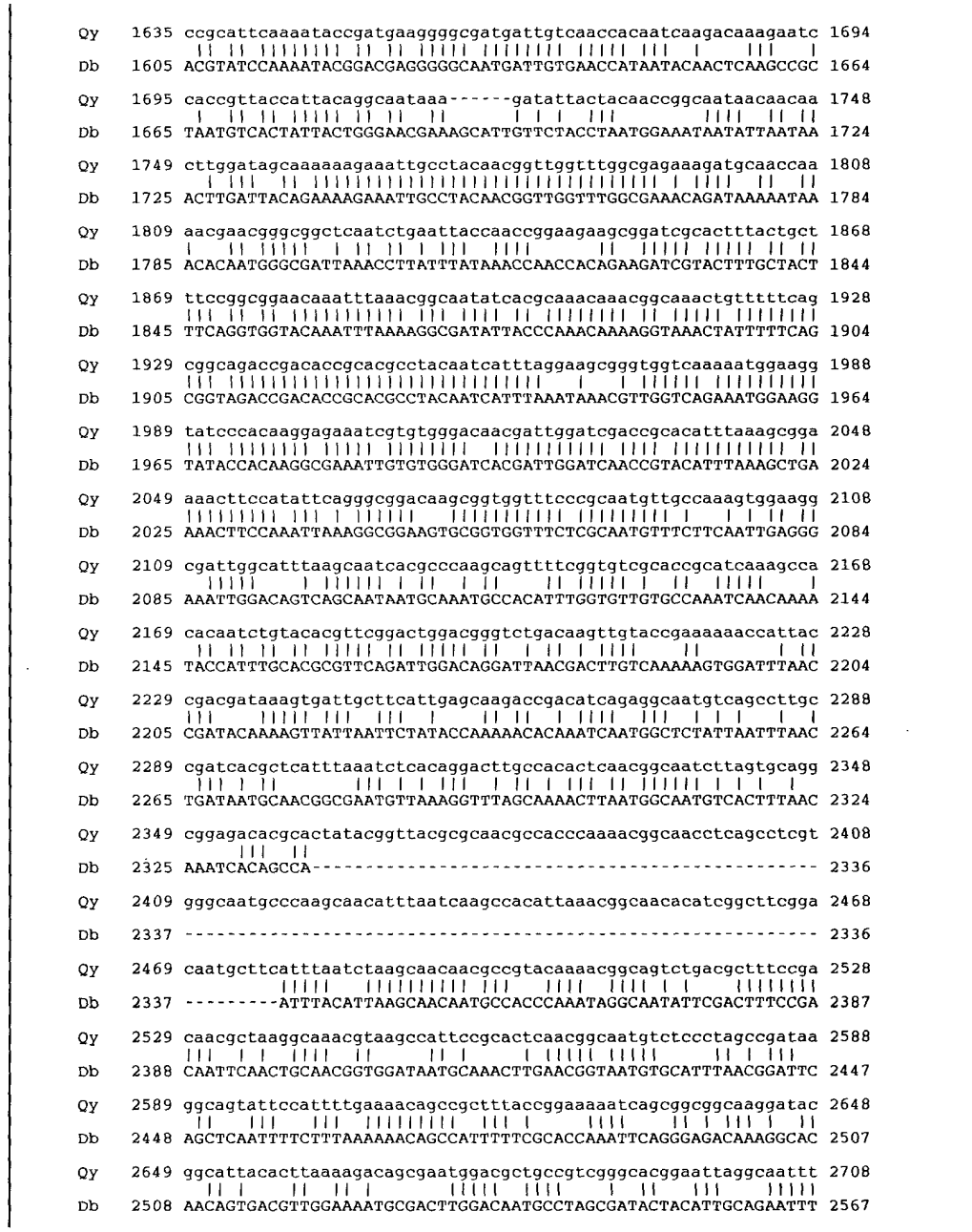
Qy     318 cgatcaatatattgtgagcgtggcacataacggcggctataacaatgttgattttggtgc 377
             |||||||||||||||||||||||||||||||  || ||||   ||||||||||||||||
Db     327 AAATCAATATATTGTGAGCGTGGCACATAACGTAGGATATACAGATGTTGATTTTGGTGC 386

Qy     378 ggagggaagcaatcccgatcagcaccgcttttcttaccaaattgtgaaaagaaataatta 437
           ||||||| ||| |||||||| || || ||| ||||  | |||||  ||| ||||||| ||
Db     387 AGAGGGAAACAACCCCGATCAACATCGTTTTACTTATAAGATTGTAAAACGAAATAACTA 446

Qy     438 taaagcagggactaacggccatccttatggcggcgattatcatatgccgcgtttgcacaa 497
           |||  ||    |||    ||||||||||  | |||||| |||| || || || || ||
Db     447 CAAAAAAG---ATAATTTACATCCTTATGAGGACGATTACCATAATCCACGATTACATAA 503

Qy     498 atttgtcacagatgcagaacctgttgagatgaccagttatatggatgggtggaaatacgc 557
           ||| || |||||| || |  ||  |||| |||||     ||||| |||| | |  ||  |
```





```
Qy   2709 aaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcggcaggcgc 2768
          ||  ||  | |||   || || ||  | |||||| || |||  |
Db   2568 AACGCTAAATAACAGTACGATCACGTTAAATTCAGCTTATTCA---------------- 2610

Qy   2769 gcaaaccggcagtgcggcagatgcgccgcgccgccgttcgcgccgttccctattatccgt 2828
               |  ||  |     || |||| ||          ||||||||  ||    | |
Db   2611 -------GCTAGCTCAAACAATACGCCACG---------TCGCCGTTCATTAGAGACGGA 2654

Qy   2829 tacgccgccaacttcggcagaatcccgtttcaacacgctgacggtaaacggcaaattgaa 2888
          ||  |||||||| |||||||||   ||||||||||  |||| |||||  || |||||||
Db   2655 AACAACGCCAACATCGGCAGAACATCGTTTCAACACATTGACAGTAAATGGTAAATTGAG 2714

Qy   2889 cggtcagggaacattccgctttatgtcggaactcttcggctaccgcagcggcaaattgaa 2948
          || || || |||||||  ||||  ||      | || |||||    ||||  |||||| ||
Db   2715 TGGGCAAGGCACATTCCAATTTACTTCATCTTTATTTGGCTATAAAAGCGATAAATTAAA 2774

Qy   2949 gctggcggaaagttccgaaggcacttacaccttggctgtcaacaataccggcaacgaacc 3008
            |  |  |     | || |||  ||||| ||  ||||   ||| || ||||| |||||
Db   2775 ATTATCCAATGACGCTGAGGGCGATTACATATTATCTGTTCGCAACACAGGCAAAGAACC 2834

Qy   3009 cgtaagtctcgagcaattgacggtagtggaaggaaaagacaacacaccgctgtccgaaaa 3068
          || ||  || |||||||| ||  | || ||| | ||||| ||   |||| | || || ||
Db   2835 CGAAACCCTTGAGCAATTAACTTTGGTTGAAAGCAAAGATAATCAACCGTTATCAGATAA 2894

Qy   3069 tcttaatttcaccctgcaaaacgaacacgtcgatgccggcgcatggcgttatcagcttat 3128
          || || || ||  |  |||| || ||||| ||||| || ||||  |||||| |  |  |
Db   2895 GCTCAAATTTACTTTAGAAAATGACCACGTTGATGCAGGTGCATTACGTTATAAATTAGT 2954

Qy   3129 ccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaagagctttccgacaa 3188
            || || ||||| |||||| ||||||| ||  | ||||| || ||  |    | | |
Db   2955 GAAGAATGATGGCGAATTCCGCTTGCATAACCCAATAAAAGAGCAGGAATTGCACAATGA 3014

Qy   3189 actcggcaaggcgggagaaacagaggccgccttgacggcaaaacaggcacaacttgccgc 3248
           | |  |  || |  || ||||     | ||    || ||||| |   |||   | ||
Db   3015 TTTAGTAAGAGCAGAGCAAGCAGAACGAACATTAGAAGCCAAACAAGTTGAACCGACTGC 3074

Qy   3249 caaacaacaggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcggccgggcg 3308
          |||  |||  | |   |    || | ||          ||   ||  || ||      |
Db   3075 TAAAACACAAACAGGTGAGCCAAAAGTGCGGTCAAGAAGAGCAGCGAGAGCAGCGTTTCC 3134

Qy   3309 caatgccaccgaaaaggcagaaagtgttgccgaaccggcccggcaggcaggcggggaaaa 3368
            || ||       |  |       |  || |    |      | ||  |  |      |
Db   3135 TGATACCCTGCCTGATCAAAGCCTGTTAAACGCATTAGAAGCCAAACAAGCTGAACTGAC 3194

Qy   3369 tgccggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaagacaccgc 3428
          |||  |  |      |    ||| | | | ||||   |||| | |  | | ||
Db   3195 TGCTGAAACACAAAAAAGTAAGGCAAAAACAAAAAAAGTGCGGTCAAAAAGAGCAGTGTT 3254

Qy   3429 cttggcgaaacagcgcgaagcggaaacccggccggctaccaccgccttcccccgcgcccg 3488
           |  |     | ||  ||          |         ||||  |
Db   3255 TTCTGATCCCCTGCTTGATCAAAGCCTGTTCGCATTAGAAGCCGCACTTGAGGTTATTGA 3314

Qy   3489 ccgcgcccgccgggatttgccgcaaccgcagccccaaccgcaaccccaaccgcagcgc-- 3546
             | | |  |           | |  |   | |||    || |  ||   || |||
Db   3315 TGCCCCACAGCAATCGGAAAAAGATCGTCTAGCTCAAGAAGAAGCGGAAAAACAACGCAA 3374

Qy   3547 -------gacctgatcagccgttatgccaatagcggtttgagtgaattttccgccacgct 3599
                  ||| ||||||||||||||| | ||||| |  ||    ||||| || || ||  |
Db   3375 ACAAAAAGACTTGATCAGCCGTTATTCAAATAGTGCGTTATCAGAATTATCTGCAACAGT 3434

Qy   3600 caacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagaccgccgcaa 3659
          || ||  |  |  | || || || ||||| || ||  | ||||  ||  |    |
Db   3435 AAATAGTATGCTTTCTGTTCAAGATGAATTAGATCGTCTTTTTGTAGATCAAGCACAATC 3494

Qy   3660 cgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagatttccgcgc 3719
          |||||  |||||||     |   | ||| |  | || |||   |||  | |  ||||| ||
Db   3495 TGCCGTGTGGACAAATATCGCACAGGATAAAAGACGCTATGATTCTGATGCGTTCCGTGC 3554

Qy   3720 ctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggcagcgggcg 3779
          || |  ||  ||||  || | || ||||| ||  |||| |||  | | |  |  || ||
Db   3555 TTATCAGCAGAAAACGAACTTACGTCAAATTGGGGTGCAAAAAGCCTTAGCTAATGGACG 3614

Qy   3780 cgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggcatcggcaa 3839
```

```
               | ||     | || |||||| | |||  | |  || ||||| || ||     |   ||
Db   3615 AATTGGGGCAGTTTTCTCGCATAGCCGTTCAGATAATACCTTTGATGAACAGGTTAAAAA 3674

Qy   3840 ctcggcacggcttgcccacggtgccgttttcgggcaatacggcatcggcaggttcgacat 3899
             ||     |  |          | | ||| |  |||||       |||   ||  |  |
Db   3675 TCACGCGACATTAACGATGATGTCGGGTTTTGCCCAATATCAATGGGGCGATTTACAATT 3734

Qy   3900 cggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatcagaggcaa 3959
          ||  | | || |||  |||| | |||    | |   |  | || |     ||  | ||
Db   3735 TGGTGTAAACGTGGGAACGGGAATCAGTGCGAGTAAAATGGCTGAAGAACAAAGCCGAAA 3794

Qy   3960 aatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggtttcggcgg 4019
          ||| |  ||    | | |  |||| ||| |  | ||||| || |      |||| ||
Db   3795 AATTCATCGAAAAGCGATAAATTATGGCGTGAATGCAAGTTATCAGTTCCGTTTAGGGCA 3854

Qy   4020 attcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcggattaccg 4079
          ||| |||||  | ||  |  | || |  |  ||||||||  |  ||   |  |||| |
Db   3855 ATTGGGCATTCAGCCTTATTTTGGAGTTAATCGCTATTTTATTGAACGTGAAAATTATCA 3914

Qy   4080 atacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgcgcgggcat 4139
          ||  ||  | || |   |    || ||  |||||||||| || |||||    || |||||
Db   3915 ATCTGAGGAAGTGAGAGTGAAAACGCCTAGCCTTGCATTTAATCGCTATAATGCTGGCAT 3974

Qy   4140 taaggcagattattcattcaaaccggcgcaacacatttccatcacgccttatttgagcct 4199
          |   |  |||||| |||| |  ||| |  |  | ||   | | | |||||||||   | |
Db   3975 TCGAGTTGATTATACATTTACTCCGACAGATAATATCAGCGTTAAGCCTTATTTCTTCGT 4034

Qy   4200 gtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgccgtattggc 4259
           |||  ||||   |   ||  || || | |||    || |||  | | || |||
Db   4035 CAATTATGTTGATGTTTCAAACGCTAACGTACAAACCACGGTAAATCTCACGGTGTTGCA 4094

Qy   4260 gcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatcaaaggttt 4319
           ||    || ||        |   |||  ||| |||| || |||||   |  |||
Db   4095 ACAACCATTTGGACGTTATTGGCAAAAAGAAGTGGGATTAAAGGCAGAAATTTTACATTT 4154

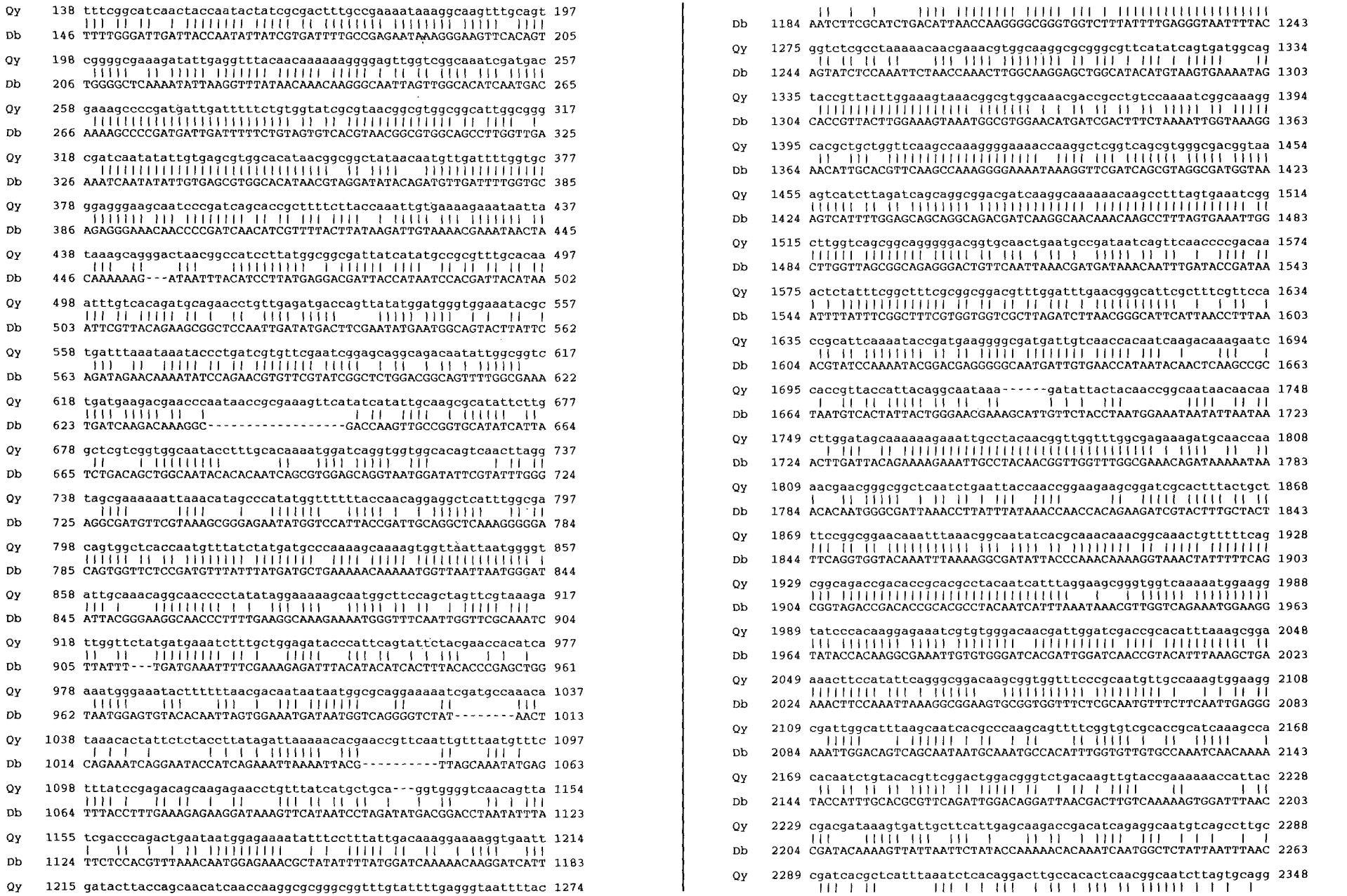
Qy   4320 cacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcagcacagcgc 4379
          |    | |||           |     |  | ||  | ||| | |     ||||| |  |
Db   4155 CCAAATTTCCGCTTTTATCTCAAAATCTCAAGGTTCACAACTCGGCAAACAGCAAAATGT 4214

Qy   4380 gggcatcaaattaggctaccgctggtaa 4407
          |||| | ||||| ||||| || ||||||
Db   4215 GGGCGTGAAATTGGGCTATCGTTGGTAA 4242


RESULT  14
AR157320
LOCUS       AR157320                4319 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6245337.
ACCESSION   AR157320
VERSION     AR157320.1  GI:16218252
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4319)
  AUTHORS   St. Geme,J.W. III and Falkow,S.
  TITLE     Haemophilus adherence and penetration proteins
  JOURNAL   Patent: US 6245337-A 1 12-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .4319
                     /organism="unknown"
BASE COUNT     1497 a    776 c    891 g   1155 t
ORIGIN


  Query Match           29.4%;  Score 1297.8;  DB 6;  Length 4319;
  Best Local Similarity   59.1%;  Pred. No. 1e-293;
  Matches 2571;  Conservative    0;  Mismatches 1567;  Indels  213;  Gaps   12;

Qy     78 ttacttagccatatgcctgtcgttcggcattctgccccaagcccgggcgggacacactta 137
          ||  ||| ||   ||| | || || || ||  |  | |||||  |||| || ||||||||
Db     86 TTTTTTAACCGCTTGCATTTCATTAGGGATAGTATCGCAAGCGTGGGCTGGTCACACTTA 145
```



```
Db   2264 TGATAATGCAACGGCGAATGTTAAAGGTTTAGCAAAACTTAATGGCAATGTCACTTTAAC 2323

Qy   2349 cggagacacgcactatacggttacgcgcaacgccacccaaaacggcaacctcagcctcgt 2408

Db   2324 AAATCACAGCCA---------------------------------------------------- 2335

Qy   2409 gggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcggcttcgga 2468

Db   2336 ---------------------------------------------------------- 2335

Qy   2469 caatgcttcatttaatctaagcaacaacgccgtacaaaacggcagtctgacgctttccga 2528

Db   2336 ---------ATTTACATTAAGCAACAATGCCACCCAAATAGGCAATATTCGACTTTCCGA 2386

Qy   2529 caacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctccctagccgataa 2588

Db   2387 CAATTCAACTGCAACGGTGGATAATGCAAACTTGAACGGTAATGTGCATTTAACGGATTC 2446

Qy   2589 ggcagtattccattttgaaaacagccgctttaccggaaaaatcagcggcggcaaggatac 2648

Db   2447 AGCTCAATTTTCTTTAAAAAACAGCCATTTTTCGCACCAAATTCAGGGAGACAAAGGCAC 2506

Qy   2649 ggcattacacttaaaagacagcgaatggacgctgccgtcgggcacggaattaggcaattt 2708

Db   2507 AACAGTGACGTTGGAAAATGCGACTTGGACAATGCCTAGCGATACTACATTGCAGAATTT 2566

Qy   2709 aaaccttgacaacgccaccattacactcaattccgcctatcgacacgatgcggcaggcgc 2768

Db   2567 AACGCTAAATAACAGTACGATCACGTTAAATTCAGCTTATTCA----------------- 2609

Qy   2769 gcaaaccggcagtgcggcagatgcgccgcgccgccgttcgcgccgttccctattatccgt 2828

Db   2610 -------GCTAGCTCAAACAATACGCCACG---------TCGCCGTTCATTAGAGACGGA 2653

Qy   2829 tacgccgccaacttcggcagaatcccgtttcaacacgctgacggtaaacggcaaattgaa 2888

Db   2654 AACAACGCCAACATCGGCAGAACATCGTTTCAACACATTGACAGTAAATGGTAAATTGAG 2713

Qy   2889 cggtcagggaacattccgctttatgtcggaactcttcggctaccgcagcggcaaattgaa 2948

Db   2714 TGGGCAAGGCACATTCCAATTTACTTCATCTTTATTTGGCTATAAAAGCGATAAATTAAA 2773

Qy   2949 gctggcggaaagttccgaaggcacttacaccttggctgtcaacaataccggcaacgaacc 3008

Db   2774 ATTATCCAATGACGCTGAGGGCGATTACATATTATCTGTTCGCAACACAGGCAAAGAACC 2833

Qy   3009 cgtaagtctcgagcaattgacggtagtggaaggaaaagacaacacaccgctgtccgaaaa 3068

Db   2834 CGAAACCCTTGAGCAATTAACTTTGGTTGAAAGCAAAGATAATCAACCGTTATCAGATAA 2893

Qy   3069 tcttaatttcaccctgcaaaacgaacacgtcgatgccggcgcatggcgttatcagcttat 3128

Db   2894 GCTCAAATTTACTTTAGAAAATGACCACGTTGATGCAGGTGCATTACGTTATAAATTAGT 2953

Qy   3129 ccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaagagctttccgacaa 3188

Db   2954 GAAGAATGATGGCGAATTCCGCTTGCATAACCCAATAAAAGAGCAGGAATTGCACAATGA 3013

Qy   3189 actcggcaaggcgggagaaacagaggccgccttgacggcaaaacaggcacaacttgccgc 3248

Db   3014 TTTAGTAAGAGCAGAGCAAGCAGAACGAACATTAGAAGCCAAACAAGTTGAACCGACTGC 3073

Qy   3249 caaacaacaggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcggccgggcg 3308

Db   3074 TAAAACACAAACAGGTGAGCCAAAAGTGCGGTCAAGAAGAGCAGCGAGAGCAGCGTTTCC 3133

Qy   3309 caatgccaccgaaaaggcagaaagtgttgccgaaccggcccggcaggcaggcggggaaaa 3368

Db   3134 TGATACCCTGCCTGATCAAAGCCTGTTAAACGCATTAGAAGCCAAACAAGCTGAACTGAC 3193

Qy   3369 tgccggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaagacaccgc 3428

Db   3194 TGCTGAAACACAAAAAAGTAAGGCAAAAACAAAAAAAGTGCGGTCAAAAAGAGCAGTGTT 3253

Qy   3429 cttggcgaaacagcgcgaagcggaaacccggccggctaccaccgccttccccgcgcccg 3488

Db   3254 TTCTGATCCCCTGCTTGATCAAAGCCTGTTCGCATTAGAAGCCGCACTTGAGGTTATTGA 3313

Qy   3489 ccgcgcccgccgggatttgccgcaaccgcagccccaaccgcaaccccaaccgcagcgc-- 3546

Db   3314 TGCCCCACAGCAATCGGAAAAAGATCGTCTAGCTCAAGAAGAAGCGGAAAAACAACGCAA 3373

Qy   3547 -------gacctgatcagccgttatgccaatagcggtttgagtgaattttccgccacgct 3599

Db   3374 ACAAAAAGACTTGATCAGCCGTTATTCAAATAGTGCGTTATCAGAATTATCTGCAACAGT 3433

Qy   3600 caacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagaccgccgcaa 3659

Db   3434 AAATAGTATGCTTTCTGTTCAAGATGAATTAGATCGTCTTTTTGTAGATCAAGCACAATC 3493

Qy   3660 cgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagatttccgcgc 3719

Db   3494 TGCCGTGTGGACAAATATCGCACAGGATAAAAGACGCTATGATTCTGATGCGTTCCGTGC 3553

Qy   3720 ctaccgccaac---aaaccgacctgcgccaaatcggtatgcagaaaaacctcggcagcgg 3776

Db   3554 TTATCAGCAGCAGAAAACGAACTTACGTCAAATTGGGGTGCAAAAAGCCTTAGCTAATGG 3613

Qy   3777 gcgcgtcggcatcctgttttcgcacaaccggaccggaaacaccttcgacgacggcatcgg 3836

Db   3614 ACGAATTGGGGCAGTTTTCTCGCATAGCCGTTCAGATAATACCTTTGATGAACAGGTTAA 3673

Qy   3837 caactcggcacggcttgcccacggtgccgtttttcgggcaatacggcatcggcaggttcga 3896

Db   3674 AAATCACGCGACATTAACGATGATGTCGGGTTTTGCCCAATATCAATGGGGCGATTTACA 3733

Qy   3897 catcggcatcagcgcgggcgcgggttttagtagcggcagcctttcagacggcatcagagg 3956

Db   3734 ATTTGGTGTAAACGTGGGAACGGGAATCAGTGCGAGTAAAATGGCTGAAGAACAAAGCCG 3793

Qy   3957 caaaatccgccgccgcgtgctgcattacggcattcaggcaagataccgcgcaggtttcgg 4016

Db   3794 AAAAATTCATCGAAAAGCGATAAATTATGGCGTGAATGCAAGTTATCAGTTCCGTTTAGG 3853

Qy   4017 cggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcggatta 4076

Db   3854 GCAATTGGGCATTCAGCCTTATTTTGGAGTTAATCGCTATTTTATTGAACGTGAAAATTA 3913

Qy   4077 ccgatacgaaaacgtcaatatcgccaccccgggccttgcattcaaccgctaccgcgcggg 4136

Db   3914 TCAATCTGAGGAAGTGAGAGTGAAAACGCCTAGCCTTGCATTTAATCGCTATAATGCTGG 3973

Qy   4137 cattaaggcagattattcattcaaaccggcgcaacacatttccatcacgccttatttgag 4196

Db   3974 CATTCGAGTTGATTATACATTTACTCCGACAGATAATATCAGCGTTAAGCCTTATTTCTT 4033

Qy   4197 cctgtcctataccgatgccgcttccggcaaagtccgaacgcgcgtcaataccgccgtatt 4256

Db   4034 CGTCAATTATGTTGATGTTTCAAACGCTAACGTACAAACCACGGTAAATCTCACGGTGTT 4093

Qy   4257 ggcgcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatcaaagg 4316

Db   4094 GCAACAACCATTTGGACGTTATTGGCAAAAAGAAGTGGGATTAAAGGCAGAAATTTTACA 4153

Qy   4317 tttcacgctgtccctccacgctgccgccgccaaggggccgcaattggaagcgcagcacag 4376

Db   4154 TTTCCAAATTTCCGCTTTTATCTCAAAATCTCAAGGTTCACAACTCGGCAAACAGCAAAA 4213

Qy   4377 cgcgggcatcaaattaggctaccgctggtaa 4407

Db   4214 TGTGGGCGTGAAATTGGGCTATCGTTGGTAA 4244


RESULT 15
NLIGA3327
LOCUS       NLIGA3327                578 bp    DNA     linear   BCT 21-MAY-1998
```

```
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3327.
ACCESSION   AJ001740
VERSION     AJ001740.1  GI:3152399
KEYWORDS    IgA protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 578)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Neisseria lactamica"
                     /isolate="NL3327"
                     /db_xref="taxon:486"
     CDS             <1. .>576
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA04965.1"
                     /db_xref="GI:3152400"
                     /db_xref="SPTREMBL:O69746"
                     /translation="GKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALV
                     GDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRL
                     HKFVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGAWLI
                     GGNTHMQGWGNNGVVSLSGDVRHANDYGPMPI"
     repeat_region   88. .95
                     /rpt_type=INVERTED
     repeat_region   139. .148
BASE COUNT      163 a    125 c    156 g    134 t
ORIGIN
```

```
  Query Match             7.5%;  Score 328.4;  DB 1;  Length 578;
  Best Local Similarity   84.7%;  Pred. No. 3.5e-66;
  Matches  381;  Conservative    0;  Mismatches   66;  Indels    3;  Gaps    1;

Qy     184 ggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtc 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 60

Qy     244 ggcaaatcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggcgtg 303
           |||||||| ||||| |||||||||||||||||||||| ||||| ||||||||||||||||
Db      61 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTG 120

Qy     304 gcggcattggcgggcgatcaatatattgtgagcgtggcacataacggcggctataacaat 363
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAAT 180

Qy     364 gttgattttggtgcggagggaagcaatcccgatcagcaccgctttttcttaccaaattgtg 423
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTTGATTTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG 240

Qy     424 aaaagaaataattataaagcagggactaacggccatccttatggcggcgattatcatatg 483
           |||||||||| || ||| | |     ||    || |||||   ||||||||| ||||||
Db     241 AAAAGAAATAACTACAAACCTG---ACAATTCACACCCTTACAACGGCGATTACCATATG 297

Qy     484 ccgcgtttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatggat 543
           |||||  |||| || || || || ||||||||||||||| || |||||  || | |||
Db     298 CCGCGCCTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGG 357

Qy     544 gggtggaaatacgctgatttaaataaataccctgatcgtgttcgaatcggagcaggcaga 603
           |||   |  || | |||  | | ||||| || || ||||| || |||||  ||||
Db     358 GGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACAC 417

Qy     604 caatattggcggtctgatgaagacgaaccc 633
           || |||||||| | |||||| ||| ||| |
Db     418 CACTATTGGCGTTATGATGATGACAAACAC 447
```

```
RESULT  16
NLIGA3354
LOCUS       NLIGA3354                584 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3354.
ACCESSION   AJ001739
VERSION     AJ001739.1  GI:3152397
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 584)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Neisseria lactamica"
                     /isolate="NL3354"
                     /db_xref="taxon:486"
     CDS             <1. .>582
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04964.1"
                     /db_xref="GI:3152398"
                     /db_xref="SPTREMBL:O69747"
                     /translation="FAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGD
                     QYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLHK
                     FVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDDKHGDLSYSGAWLIG
                     GNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGA"
     repeat_region   82. .88
                     /rpt_type=INVERTED
     repeat_region   134. .142
BASE COUNT      163 a    127 c    159 g    135 t
ORIGIN
```

```
  Query Match             7.3%;  Score 321.6;  DB 1;  Length 584;
  Best Local Similarity   84.8%;  Pred. No. 1.4e-64;
  Matches  373;  Conservative    0;  Mismatches   64;  Indels    3;  Gaps    1;

Qy     190 tttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaa 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAA 60

Qy     250 tcgatgacgaaagccccgatgattgatttttctgtggtatcgcgtaacggcgtggcggca 309
           || |||||| |||||||||||||||||||||| ||||| ||||||||||||||||||||
Db      61 TCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCA 120

Qy     310 ttggcgggcgatcaatatattgtgagcgtggcacataacggcggctataacaatgttgat 369
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGAT 180

Qy     370 tttggtgcggagggaagcaatcccgatcagcaccgcttttcttaccaaattgtgaaaaga 429
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGA 240

Qy     430 aataattataaagcagggactaacggccatccttatggcggcgattatcatatgccgcgt 489
           ||||| || ||| | |     ||    || |||||   ||||||||| ||||||||||
```

```
Db    241 AATAACTACAAACCTG---ACAATTCACACCCTTACAACGGCGATTACCATATGCCGCGC 297

Qy    490 ttgcacaaatttgtcacagatgcagaacctgttgagatgaccagttatatggatgggtgg 549
          |||| || || || || ||||||||||||||| || |||||  || | |||   |||
Db    298 CTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGGAAT 357

Qy    550 aaatacgctgatttaaataaataaccctgatcgtgttcgaatcggagcaggcagacaatat 609
          |  ||  | |||  | | ||||| || || ||||| || |||||. ||||    || |||
Db    358 ACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTAT 417

Qy    610 tggcggtctgatgaagacga 629
          ||||| | |||||| |||||
Db    418 TGGCGTTATGATGATGACGA 437


RESULT  17
NLIGA823
LOCUS       NLIGA823                572 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL823.
ACCESSION   AJ001737
VERSION     AJ001737.1  GI:3152389
KEYWORDS    IgA protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 572)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Neisseria lactamica"
                     /isolate="NL823"
                     /db_xref="taxon:486"
     CDS             <1. .304
                     /codon_start=2
                     /transl_table=11
                     /product="IgA protease homologue"
                     /protein_id="CAA04962.1"
                     /db_xref="GI:3152390"
                     /db_xref="SPTREMBL:O69748"
                     /translation="KDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGDQYIVS
                     VAHNGGYNNVDFGAEGGNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLPGISS"
     repeat_region   44. .51
                     /rpt_type=INVERTED
     repeat_region   68. .75
BASE COUNT      158 a    124 c    157 g    133 t
ORIGIN


  Query Match             6.6%;  Score 290.2;  DB 1;  Length 572;
  Best Local Similarity   82.5%;  Pred. No. 3.3e-57;
  Matches  359;  Conservative    0;  Mismatches   68;  Indels    8;  Gaps    2;

Qy    204 gaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatcgatgacgaaagc 263
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||||
Db      1 GAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 60

Qy    264 cccgatgattgatttttctgtggtatcgcgtaacggcgtggcggcattggcgggcgatca 323
          |||||||||||||||||| ||||| |||||||||||||||||||||||| |||||||||
Db     61 CCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCA 120

Qy    324 atatattgtgagcgtggcacataacggcggctataacaatgttgattttggtgcggaggg 383
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    121 ATATATTGTGAGCGTGGCGCATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGG 180

Qy    384 aagcaatcccgatcagcaccgcttttcttaccaaattgtgaaaagaaataattataaagc 443
          | ||||||||||||||||||||||||||||||||||||||||| |||||||||| || ||| |
Db    181 AGGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTAAAAAGAAATAACTACAAACC 240

Qy    444 agggactaacggccatccttatggcggcgattatcatatgccgcg-----tttgcacaaa 498
          |     ||    || |||||   |||||||| ||||||||||      | ||| ||
Db    241 TG---ACAATTCACACCCTTACAACGGCGATTACCATATGCCGCGCCTGCCTGGCATAAG 297

Qy    499 tttgtcacagatgcagaacctgttgagatgaccagttatatggatgggtggaaatacgct 558
          || || || |||||||||||||| || |||||  || | |||   |||   |  ||  |
Db    298 TTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGGAATACCTATTCC 357

Qy    559 gatttaaataaatacccgatcgtgttcgaatcggagcaggcagacaatattggcggtct 618
          |||  | | ||||| || || ||||| || |||||  ||||    || |||||||| | |
Db    358 GATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTATTGGCGTTAT 417

Qy    619 gatgaagacgaaccc 633
          ||||| ||| ||| |
Db    418 GATGATGACAAACAC 432


RESULT  18
NLIGA3293
LOCUS       NLIGA3293               572 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3293.
ACCESSION   AJ001738
VERSION     AJ001738.1  GI:3152385
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 572)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Neisseria lactamica"
                     /isolate="NL3293"
                     /db_xref="taxon:486"
     CDS             <1. .>570
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04963.1"
                     /db_xref="GI:3152386"
                     /db_xref="SPTREMBL:O69745"
                     /translation="KDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGDQYIVS
                     VAHNVGYRDVDFGAEGSNPDQHRFSYKIAKRNNYKNDETHPYEKDYHNPRLHKFVTEA
                     APIDMTSDMDGNKYTDRTKYPERVRIGSGWQFWRNDQDKGDQVAGAYHYLTAGNTHNQ
                     GGAGGGWSSLSGDVRHAGNYGPIPIAAGSS"
     repeat_region   67. .74
                     /rpt_type=INVERTED
     repeat_region   118. .127
BASE COUNT      169 a    119 c    154 g    130 t
ORIGIN


  Query Match             6.2%;  Score 275.4;  DB 1;  Length 572;
  Best Local Similarity   77.0%;  Pred. No. 9.9e-54;
  Matches  349;  Conservative    0;  Mismatches  101;  Indels    3;  Gaps    1;

Qy    205 aaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatcgatgacgaaagcc 264
          |||||||||||||||||||||||||||||||||||||||||| ||||| ||||| ||||||
```

```
Db      1 AAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGAAAATCAATGACAAAAGCC 60

Qy    265 ccgatgattgatttttctgtggtatcgcgtaacggcgtggcggcattggcgggcgatcaa 324
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     61 CCGATGATTGATTTTTCTGTGGTATCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAA 120

Qy    325 tatattgtgagcgtggcacataacggcggctataacaatgttgattttggtgcggaggga 384
          |||||||||||||||| ||||| |  || |||  | ||||||||||||||||||||||
Db    121 TATATTGTGAGCGTGGCGCATAATGTAGGTTATCGCGATGTTGATTTTGGTGCGGAGGGA 180

Qy    385 agcaatcccgatcagcaccgcttttcttaccaaattgtgaaaagaaataattataaagca 444
          |||||||||||||||||| ||||||||  | ||||  ||| | ||||||||||
Db    181 AGCAATCCCGATCAGCACCGTTTTTCTTATAAGATTGCAAAACGGAATAATTATAAAAAC 240

Qy    445 gggactaacggccatccttatggcggcgattatcatatgccgcgtttgcacaaatttgtc 504
          |    | |   ||||||||||     |||||| || |  || || || || ||||||||
Db    241 G---ATGAAACGCATCCTTATGAGAAAGATTACCACAACCCACGCTTACATAAATTTGTT 297

Qy    505 acagatgcagaacctgttgagatgaccagttatatggatgggtggaaatacgctgattta 564
          || || || |  ||  |||| |||||   | |||||||| ||    |||||  | |||
Db    298 ACGGAAGCCGCCCCGATTGATATGACTTCTGATATGGACGGCAACAAATATACAGATAGG 357

Qy    565 aataaataccctgatcgtgttcgaatcggagcaggcagacaatattggcggtctgatgaa 624
          |  ||||| || || || || || |||||  | ||  | || | ||||||    ||| ||
Db    358 ACGAAATATCCCGAACGCGTCCGTATCGGCTCCGGGTGGCAGTTTTGGCGAAACGATCAA 417

Qy    625 gacgaacccaataaccgcgaaagttcatatcat 657
          ||| ||  | || |    |   |  ||||||||
Db    418 GACAAAGGCGATCAAGTTGCCGGCGCATATCAT 450


RESULT 19
AF012207
LOCUS       AF012207               2992 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z3524 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012207
VERSION     AF012207.1  GI:2529497
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2992
                     /organism="Neisseria meningitidis"
                     /strain="Z3524"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup III; isolated in Chad, 1988"
     mat_peptide     1. .>2992
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2992
                     /gene="iga"
                     /allele="iga3"
     CDS             <1. .>2992
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45790.2"
                     /db_xref="GI:12597967"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     GANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGANNDITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
                     GIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
                     RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGQTENNDWVFMGYKQEEAQKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAR
                     KDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVSNITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFGATQINGNVNLSQNAALVLGKAALW
                     GQIQGQGNSSVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKINH
                     LSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQDR
                     SHLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQKA
                     TQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      986 a    685 c    640 g    681 t
ORIGIN


  Query Match             4.4%;  Score 194.8;  DB 1;  Length 2992;
  Best Local Similarity   59.0%;  Pred. No. 1.1e-34;
  Matches  353;  Conservative    0;  Mismatches  242;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          |||  || ||    ||| || |       |||||||||||||||||||||||| ||| ||
Db   1036 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1095

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     |  ||| || ||    || |||| ||| ||||
Db   1096 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGGA 1155

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1156 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1215

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1216 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1275

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db   1276 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1335

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || |||||  | ||||||||  | || ||  ||  | | |||      |
Db   1336 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1395

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          |||||  | || |||||  ||  | |||||||| || || ||||| |||||||||   || |
Db   1396 ATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1455

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1456 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1515

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   ||||   |  ||||  | ||  | || |||||   |   |||||  |  |
Db   1516 ACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1575

Qy   1736 gcaataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcga 1793
          |||         |  || | |  ||| ||    |   | ||| ||  | |||| |
Db   1576 AAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTA 1633


RESULT 20
AE002424/c
LOCUS       AE002424               10485 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 66 of 206 of
```

```
            the complete genome.
ACCESSION   AE002424 AE002098
VERSION     AE002424.1  GI:7225923
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10485)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10485)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10485
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            complement(67. .795)
                     /gene="NMB0696"
     CDS             complement(67. .795)
                     /gene="NMB0696"
                     /note="similar to GB:M61017 SP:P27675 PID:142988 percent
                     identity: 81.40; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="amino acid ABC transporter, ATP-binding protein"
                     /protein_id="AAF41113.1"
                     /db_xref="GI:7225924"
                     /translation="MIKFKNVHKHFKDLHVINGVNLEIKKGEVVVVCGPSGSGKSTLI
                     RTVNQLESIESGEIWVDGVNVADPKTDLNKIREEVGFVFQGFNLYPHLTVLENITLAP
                     MKVKGQNAEQAEKKAMELLERVGLAHKKDAFPSQLSGGQQQRVAIARGLAMEPRVMLF
                     DEPTSALDPEMVGEVLKVMKDLAESGMTMMCVTHEMGFAREVADRVIFVDKGQILEDE
                     TPEAFFTNPKHERAKQFLQQVMTH"
     gene            complement(968. .1747)
                     /gene="NMB0697"
     CDS             complement(968. .1747)
                     /gene="NMB0697"
                     /note="similar to GB:D10483 SP:P06992 GB:M11054 GB:M68521
                     GB:X04711 percent identity: 69.41; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dimethyladenosine transferase"
                     /protein_id="AAF41114.1"
                     /db_xref="GI:7225925"
                     /translation="MKEHKARKRFGQNFLQDTRIISDIVNAVRPQADDVVIEIGPGLA
                     AITEPLAKKLNRLHVVEIDRDIVCRLKTLPFADKLVIHEGDVLQFDFNGIAGKKKIVG
                     NLPYNISTPLLFKLAEVADDVVDMHFMLQKEVVERMVAAPKSNDYGRLGVMLQYFFDM
                     EMLIDVPPESFDPAPKVDSAVVRMIPVKHRIGKADDFEHFAKLVKLAFHQRRKTIRNN
                     LKELAGDDDLQAVGINPQDRAEHIAPEKYVALSNYLAGKVV"
     gene            1944. .2633
                     /gene="NMB0698"
     CDS             1944. .2633
                     /gene="NMB0698"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41115.1"
                     /db_xref="GI:7225926"
                     /translation="MLFEGLPCLAVWCKNRLSAVLPVGGLNVSENLLEIETHPFDPVL
                     PPKAAVMMMGTFPPKEDKRAMQFHYPNFQNDMWRVYGLVFFNDAAHFQRLSEKAFDAE
                     KIKAFLHERGIASCPTVLKAVRQHGNASDKFLKVVETVDLAAVLAKIPECRHICTTGG
                     KATEILLDIQGGGIKMPKTGETVPFPFAGRDLTLTRLPSTSRAYPLSLAKKAAAYRAF
                     FEMAGLCEKQL"
     gene            2716. .3918
                     /gene="NMB0699"
     CDS             2716. .3918
                     /gene="NMB0699"
                     /note="similar to GB:M58444 SP:P16706 PID:141798
                     PID:461373 percent identity: 80.71; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tryptophan synthase, beta subunit"
                     /protein_id="AAF41116.1"
                     /db_xref="GI:7225927"
                     /translation="MKNYHAPDEKGFFGEHGGLYVSETLIPALQELADAYKAAKNDPE
                     FWEAFRHDLKHYVGRPSPVYHAARLSEHLGGAQIWLKREDLNHTGAHKVNNTIGQALL
                     AKRMGKKRVIAETGAGQHGVASATVAARFGMTCDVYMGADDIQRQMPNVFRMKLLGAN
                     VVGVESGSRTLKDAMNEAMREWVARVDDTFYIIGTAAGPAPYPEMVRDFQCVIGNEAK
                     AQMQEAIGRQPDVAVACVGGGSNAIGLFHPYIGEENVRLVGVEAGGLGVNTPDHAAPI
                     TSGAPIGVLHGFRSYLMQDENGQVLGTHSVSAGLDYPGIGPEHSHLHDIKRVEYTVAK
                     DDEALEAFDLLCRFEGIIPALESSHAVAWAVKNAPKMGKDQVILVNLSGRGDKDINTV
                     AKLKGIKL"
     gene            complement(4097. .9544)
                     /gene="NMB0700"
     CDS             complement(4097. .9544)
                     /gene="NMB0700"
                     /note="similar to GP:732874 percent identity: 93.40;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA-specific serine endopeptidase"
                     /protein_id="AAF41117.1"
                     /db_xref="GI:7225928"
                     /translation="MKTKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHV
                     KGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFV
                     TEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGT
                     PYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDK
                     QEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHHWKTTGT
                     NSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGINN
                     DITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVIL
                     NQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEH
                     IRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYIQNNDDDDAGYYYYRPRKP
                     IPQGKDLYFKNYRYYALKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNN
                     QRISGFSGFFGEENGKGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSG
                     RPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVS
                     DITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNV
                     NLNQNAALVLGKAALWGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNA
                     SDAQSANKYHTLKINHLSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEP
                     NQEGLNLFDASSVQDRSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVK
                     PAPSPATNTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEAKRQQAKAEQVKRQ
                     QAEAERKSAELAKQKAEAEREARELATRQKAEQERSSAELARRHEKEREAAELSAKQK
                     VEAEREAQALAVRRKAEAEEAKRQAAELARRHEKEREAAELSAKQRVGEEERRQTAQS
                     QPQRRKRRAAPQDYMAASQDRPKRRGHRSVQQNNVEIAQAQAELARRQQEERKAAELL
                     AKQRAEAEREAQALAARRKAEAEEAKRQAAELAHRQEAERKAAELSANQKAAAEAQAL
                     AARQQKALARQQEEARKAAELAVKQKAETERKTAELAKQRAAAEAAKRQQEARQTAEL
```

```
                     ARRQEAERQAAELSAKQKAETDREAAESAKRKAEEEEHRQAAQSQPQRRKRRAAPQDY
                     MAASQNRPKRRGRRSTLPAPPSPSFDSSAYAAPRALHNPDWYENDYEEIPLDALEDEN
                     VSESVDTSDKQPQDNTELHEKYENDYEEIPLDALEDEDVSESVDTSDKQPQDNTELHE
                     KVETVSLQPRAAQPRAQAATQLQAQAAAQADAVSTNTNSALSDAMASTQSILLDTGAS
                     LTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRSLSENMQI
                     GGVLTYSDSQHTFDQASGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSRLQTQQKA
                     NFNRASIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGNDSVKVSSMSVKTLT
                     AGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGNSFVYKADNQQQYSAGAALLY
                     RNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
     gene            complement(9703. .9957)
                     /gene="NMB0701"
     CDS             complement(9703. .9957)
                     /gene="NMB0701"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41118.1"
                     /db_xref="GI:7225929"
                     /translation="MTNPSVRKPAPRHSHEPPSRHYHESGNLGLRVKKTYIPSFPQKQ
                     KTKIRNLKSRHSRKSGNPKTKSRRNLTEEPNPENRLPAFA"
BASE COUNT     2182 a   2461 c   2686 g   3156 t
ORIGIN


  Query Match             4.4%;  Score 192.4;  DB 1;  Length 10485;
  Best Local Similarity   58.1%;  Pred. No. 5.2e-34;
  Matches  359;  Conservative    0;  Mismatches  256;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db   8463 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 8404

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| ||||| 
Db   8403 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 8344

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |   || ||||  |  ||   |||| |||| ||       |     |||| |
Db   8343 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 8284

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   8283 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 8224

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db   8223 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 8164

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||  | | |||     |
Db   8163 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 8104

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db   8103 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 8044

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   8043 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 7984

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   || | | |  ||||| | ||  | || || ||       |||||     ||
Db   7983 ACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGATTACCGATCCCA 7924

Qy   1736 gcaataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgaga 1795
            | | |   |      |||  ||||  |  ||  || | | | | | | |  | |
Db   7923 AAACCATCTCTATTCATTATATTCAAAATAATGATGATGACGATGCTGGTTATTACTATT 7864

Qy   1796 aagatgcaaccaaaacga 1813
          |     | |  ||| |||
Db   7863 ACCGCCCTAGGAAACCGA 7846
```

```
RESULT  21
AX044030/c
LOCUS       AX044030               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 109 from Patent WO0066791.
ACCESSION   AX044030
VERSION     AX044030.1  GI:11342914
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 109 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT     82523 a  82940 c  96712 g  87805 t
ORIGIN


  Query Match             4.4%;  Score 192.4;  DB 6;  Length 349980;
  Best Local Similarity   58.1%;  Pred. No. 9.7e-34;
  Matches  359;  Conservative    0;  Mismatches  256;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db 130037 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 129978

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| ||||| 
Db 129977 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 129918

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |   || ||||  |  ||   |||| |||| ||       |     |||| |
Db 129917 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 129858

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db 129857 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 129798

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db 129797 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 129738

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||  | | |||     |
Db 129737 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 129678

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db 129677 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 129618

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db 129617 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 129558

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   || | | |  ||||| | ||  | || || ||       |||||     ||
```

```
Db 129557 ACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGATTACCGATCCCA 129498

Qy   1736 gcaataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcgaga 1795
          |  | |   |     |||   ||||  |   ||   || | | | | | | |  | |
Db 129497 AAACCATCTCTATTCATTATATTCAAAATAATGATGATGACGATGCTGGTTATTACTATT 129438

Qy   1796 aagatgcaaccaaaacga 1813
          |     | |  ||| |||
Db 129437 ACCGCCCTAGGAAACCGA 129420


RESULT  22
NMIGA476
LOCUS       NMIGA476               1492 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NG44/76).
ACCESSION   X82481
VERSION     X82481.1  GI:732863
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1492)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1492)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1492
                     /organism="Neisseria meningitidis"
                     /strain="NG44/76"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57864.1"
                     /db_xref="GI:732864"
                     /db_xref="SPTREMBL:Q51165"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHH
                     WNITFGTNSKIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDY
                     TVKGINNDITWLGAGIDVTDGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKV
                     GDGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANG
                     NDLTFEHIRNVDEGARIVNHNTSHASTITLTGKSLITNPNSLSVHS"
BASE COUNT      500 a    317 c    322 g    353 t
ORIGIN


  Query Match             4.3%;  Score 190;  DB 1;  Length 1492;
  Best Local Similarity   61.0%;  Pred. No. 1.3e-33;
  Matches  326;  Conservative    0;  Mismatches  205;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          |||  || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db    929 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 988

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| |||||
Db    989 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1048

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || | |  || ||||  |  ||   |||| |||| ||      |    |||| |
Db   1049 TTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1108

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |    |  | | || ||| |  ||||||||
Db   1109 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1168

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db   1169 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1228

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || |||||  | ||||||||  | || ||  ||  | | |||      |
Db   1229 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1288

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          |||||  | || |||||  || || |||||||| || || |||||  |||||||||   || |
Db   1289 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1348

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    ||||||||  |||   ||||||||||
Db   1349 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1408

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattacta 1729
          ||||   |  | | |  |||||  | ||  | || || ||       ||||| |
Db   1409 ACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTAATTACCA 1462


RESULT  23
AF012208
LOCUS       AF012208               2983 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4400 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012208
VERSION     AF012208.1  GI:2529499
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2983)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2983)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2983
                     /organism="Neisseria meningitidis"
                     /strain="Z4400"
                     /db_xref="taxon:487"
                     /note="serogroup C, ET-37 complex; isolated in Mali, 1990"
     mat_peptide     1. .>2983
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2983
                     /gene="iga"
                     /allele="iga5"
     CDS             <1. .>2983
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
```

```
                     /protein_id="AAC45791.2"
                     /db_xref="GI:12597968"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHHWNITFGTNSKIGSTAVRLAGNE
                     KDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGINNDITWLGAGIDVTDGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     SHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     RKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPVPSPATNTASQAQT
                     DSAQIAKPQNIVVAPPSPQAN"
BASE COUNT      978 a    679 c    639 g    687 t
ORIGIN


  Query Match             4.3%;  Score 190;  DB 1;  Length 2983;
  Best Local Similarity   61.0%;  Pred. No. 1.5e-33;
  Matches  326;  Conservative    0;  Mismatches  205;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||| || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db   1039 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1098

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| |||||
Db   1099 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1158

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || | |  || |||  |  ||   |||| |||| ||      |     |||| |
Db   1159 TTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1218

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||||  |     |  | || ||| |  ||||||||
Db   1219 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1278

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || |||||||||  || ||  |  |||| || || || |   |     |||  |
Db   1279 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1338

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | |||||||  | || ||  ||  | | |||      |
Db   1339 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1398

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || |||| || || |||||||| || || |||| |||||||||   || |
Db   1399 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1458

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1459 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1518

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattacta 1729
          ||||   |  | | |  ||||  | ||  | || || ||       ||||| |
Db   1519 ACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTAATTACCA 1572


RESULT  24
AF012209
LOCUS       AF012209               2995 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4081 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012209
VERSION     AF012209.1  GI:2529501
KEYWORDS    .

SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Neisseria meningitidis"
                     /strain="Z4081"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup V; isolated in China, 1979;
                     also present in strain Z4063, serogroup A, subgroup VII,
                     isolated in China, 1979"
     mat_peptide     1. .>2995
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2995
                     /gene="iga"
                     /allele="iga9"
     CDS             <1. .>2995
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45792.2"
                     /db_xref="GI:12597969"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHHWNITFGTNSKIGSTAVRLAGNE
                     KDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGINNDITWLGAGIDVTDGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     SHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     RKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQK
                     ATQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      984 a    683 c    642 g    686 t
ORIGIN


  Query Match             4.3%;  Score 190;  DB 1;  Length 2995;
  Best Local Similarity   61.0%;  Pred. No. 1.5e-33;
  Matches  326;  Conservative    0;  Mismatches  205;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||| || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db   1039 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1098

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| |||||
Db   1099 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1158

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
```

```
Db   1159 TTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1218

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435

Db   1219 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1278

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495

Db   1279 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1338

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggggacggtgcaactgaatgccgata 1555

Db   1339 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1398

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615

Db   1399 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1458

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675

Db   1459 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1518

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattacta 1729

Db   1519 ACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTAATTACCA 1572


RESULT 25
AF235032
LOCUS       AF235032                4659 bp    DNA     linear   BCT 02-MAR-2001
DEFINITION  Neisseria meningitidis IgA1 protease gene, complete cds.
ACCESSION   AF235032
VERSION     AF235032.1  GI:13183004
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4659)
  AUTHORS   Vitovski,S. and Sayers,J.R.
  TITLE     Degenerate specificity of Neisseria meningitidis IgA1 protease
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4659)
  AUTHORS   Vitovski,S. and Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Division MGM, University of Sheffield, F
            floor, Royal Hallamshire Hospital, Sheffield S10 2JF, UK
FEATURES             Location/Qualifiers
     source          1. .4659
                     /organism="Neisseria meningitidis"
                     /strain="NMB"
                     /db_xref="taxon:487"
     CDS             1. .4659
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="AAK15023.1"
                     /db_xref="GI:13183005"
                     /translation="MKTKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHV
                     KGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFV
                     TEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGT
                     PYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDK
                     QEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADKIRQRDNAGTIKGNGEHHWNITFG
                     TNSKIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGIN
                     NDITWLGAGIDVTDGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVI
                     LNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFE
                     HIRNVDEGARIVNHNTSHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRR
                     PIPQGKDLYYKNYRYYALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKN
                     NQRISGFSGFFGEENGKGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLS
                     GRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNV
                     SNITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGN
                     VNLNQNAALVLGKAALWGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNN
                     ASDAQSANQYHTLKINHLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGE
                     PNQEGLNLFDASSVQDRSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRV
                     KPVPSPATNTASQAQTDSAQIAKPQNIVVAPPSPQANQAEEAKRQQAKAEQVKRQQAE
                     AERKSAELAKQKAEAERERARELATRQKAEQERSSAELARRHEKEREAAELSAKQKVEA
                     EREAQALAVRRKAEAEEAKRQAAELARQQEEARKAAELAAKQKAETERKAAEIAEQKA
                     EAEREAAELAKQKAEEEGRQAAQSQPKRRNRRAIPPELSSDATTRALPRIARNSNPDA
                     SDYEKIPLDALEDEDVSESVDTSDKQPQDNTELHEKVETVSLQPRAAQPRAQAAAQAD
                     AVSTNTNSALSDAMASTQSILLDTGASLTRHIAQKSRADAEKNSVWMSNIGYGRDYAS
                     AQYRRFSSKRTQTQIGIDRSLSENMQIGGVLTYSDSQHTFDQASGKNTFVQANLYGKY
                     YLNDAWYVAGDIGAGSLRSRLQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIR
                     YSRLSSADYKLGNDSVKVSSMSVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKG
                     GVNVGGNSFAYKADNQQQYSAGAALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF
                     "
BASE COUNT     1542 a   1074 c   1041 g   1002 t
ORIGIN


  Query Match             4.3%;  Score 190;  DB 1;  Length 4659;
  Best Local Similarity   61.0%;  Pred. No. 1.6e-33;
  Matches  326;  Conservative    0;  Mismatches  205;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258

Db   1085 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1144

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315

Db   1145 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1204

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375

Db   1205 TTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAACGGCGACAGAT 1264

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435

Db   1265 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1324

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495

Db   1325 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1384

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggggacggtgcaactgaatgccgata 1555

Db   1385 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1444

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615

Db   1445 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1504

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675

Db   1505 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1564

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattacta 1729

Db   1565 ACAACACAAGCCACGCCTCCACGATCACGCTAACGGGTAAATCTTTAATTACCA 1618


RESULT 26
NMIGA771
LOCUS       NMIGA771                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (EN3771).
ACCESSION   X82468
VERSION     X82468.1  GI:732865
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
```

```
          in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
          influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
          The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
          DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="EN3771"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57851.1"
                     /db_xref="GI:732866"
                     /db_xref="SPTREMBL:Q57309"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYT
                     VKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI"
BASE COUNT      492 a    321 c    325 g    351 t
ORIGIN
```

```
  Query Match             4.3%;  Score 189.8;  DB 1;  Length 1489;
  Best Local Similarity   60.6%;  Pred. No. 1.5e-33;
  Matches  330;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          |||  || ||    ||| || |       ||||||||||||||||.||||||| ||| ||
Db    926 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 985

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     |  ||| || ||    || |||| ||| |||||
Db    986 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGGA 1045

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |    |||| |
Db   1046 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1105

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1106 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1165

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||  |
Db   1166 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1225

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||  | | |||      |
Db   1226 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1285

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| ||  | |||||||| || || ||||| |||||||||   || |
Db   1286 ATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1345

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1346 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1405

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   ||||   |  ||||| | ||  | || |||||   |   |||||  |  |
Db   1406 ACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1465

Qy   1736 gcaat 1740
            |||
Db   1466 AAAAT 1470
```

```
RESULT  27
NMIGATH2
LOCUS       NMIGATH2                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (ETH2).
ACCESSION   X82469
VERSION     X82469.1  GI:732887
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="ETH2"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57852.1"
                     /db_xref="GI:732888"
                     /db_xref="SPTREMBL:Q57309"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYT
                     VKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI"
BASE COUNT      492 a    321 c    325 g    351 t
ORIGIN
```

```
  Query Match             4.3%;  Score 189.8;  DB 1;  Length 1489;
  Best Local Similarity   60.6%;  Pred. No. 1.5e-33;
  Matches  330;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          |||  || ||    ||| || |       ||||||||||||||||||||||| ||| ||
Db    926 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 985

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     |  ||| || ||    || |||| ||| |||||
Db    986 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGGA 1045
```

```
Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1046 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1105

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||||  |      |  | || ||| |  ||||||||
Db   1106 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1165

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db   1166 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1225

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||  | | |||      |
Db   1226 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1285

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| ||  | ||||||||| || || ||||| |||||||||   || |
Db   1286 ATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1345

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1346 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1405

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   ||||   |  ||||| | ||  | || |||||   |   |||||  |  |
Db   1406 ACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1465

Qy   1736 gcaat 1740
            |||
Db   1466 AAAAT 1470
```

RESULT 28
NMIGAG117
LOCUS NMIGAG117 1525 bp DNA linear BCT 14-MAR-1995
DEFINITION N.meningitidis iga gene (NG117).
ACCESSION X82483
VERSION X82483.1 GI:732881
KEYWORDS iga gene; IgA1 protease.
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 1525)
AUTHORS Lomholt,H., Poulsen,K. and Kilian,M.
TITLE Comparative characterization of the iga gene encoding IgA1 protease in Neisseria meningitidis, Neisseria *gonorrhoeae* and Haemophilus influenzae
JOURNAL Mol. Microbiol. 15 (3), 495-506 (1995)
MEDLINE 95302961
REFERENCE 2 (bases 1 to 1525)
AUTHORS Lomholt,H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology, The Bartholin Building, University of Aarhus, 8000 C Aarhus C, DENMARK
FEATURES Location/Qualifiers
source 1. .1525
/organism="Neisseria meningitidis"
/strain="NG117"
/db_xref="taxon:487"
gene 1. .1525
/gene="iga"
CDS <1. .>1525
/gene="iga"
/codon_start=1
/transl_table=11
/product="IgA1 protease"
/protein_id="CAA57866.1"
/db_xref="GI:732882"
/db_xref="SPTREMBL:Q51171"
/translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYA
VSVKHAKAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNM
ARFNKFVTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKG
YDLRDLSQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALT
NYGVLGDSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDN
AGTIKGNREHHWNITFGTNSKIGSTAVRLANNERDANNGQNVTFEDNGTLVLDQNINQ
GAGGLFFKGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEIN
GTGVNQGQLKVGDGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYF
GFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI
"
BASE COUNT 507 a 318 c 334 g 366 t
ORIGIN

Query Match 4.3%; Score 189.8; DB 1; Length 1525;
Best Local Similarity 60.6%; Pred. No. 1.5e-33;
Matches 330; Conservative 0; Mismatches 212; Indels 3; Gaps 1;

```
Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          |||  || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db    962 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1021

Qy   1259 ttgagggtaatttttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  ||| || |||     |  ||| || ||    || |||| ||| |||||
Db   1022 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGGA 1081

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1082 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1141

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||||  |      |  | || ||| |  ||||||||
Db   1142 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1201

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db   1202 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1261

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||  | | |||      |
Db   1262 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1321

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| ||  | ||||||||| || || ||||| |||||||||   || |
Db   1322 ATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1381

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1382 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1441

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   ||||   |  ||||| | ||  | || |||||   |   |||||  |  |
Db   1442 ACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1501

Qy   1736 gcaat 1740
            |||
Db   1502 AAAAT 1506
```

RESULT 29
NGIGAPRO
LOCUS NGIGAPRO 4899 bp DNA linear BCT 21-MAR-1995
DEFINITION Neisseria gonorrhoeae gene for IgA protease.
ACCESSION X04835
VERSION X04835.1 GI:44868
KEYWORDS IgA *protease*; inverted repeat; protease; secreted protein; signal peptide.
SOURCE Neisseria *gonorrhoeae*.
ORGANISM Neisseria gonorrhoeae
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 4899)
AUTHORS Pohlner,J., Halter,R., Beyreuther,K. and Meyer,T.F.

```
  TITLE     Gene structure and extracellular secretion of Neisseria gonorrhoeae
            IgA protease
  JOURNAL   Nature 325 (6103), 458-462 (1987)
  MEDLINE   87115823
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="Neisseria gonorrhoeae"
                     /strain="M11"
                     /db_xref="taxon:485"
                     /clone="pIP100"
     promoter        34. .39
                     /note="put. -43 region"
     promoter        43. .48
                     /note="put. -35 region"
     promoter        63. .68
                     /note="put. -10 region"
     RBS             87. .97
                     /note="put. rRNA binding region"
     CDS             104. .4702
                     /note="IgA protease precursor; 169K (AA -27 to 1505)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA28538.1"
                     /db_xref="GI:44869"
                     /db_xref="SWISS-PROT:P09790"
                     /translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
                     KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
                     VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
                     SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
                     DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
                     NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
                     KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
                     QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
                     DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
                     DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
                     ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQKRFLLTGGANLNGKI
                     SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
                     SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
                     SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
                     LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
                     QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
                     NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
                     QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
                     ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
                     IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
                     STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
                     LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
                     LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
                     LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
                     MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
                     AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
     sig_peptide     104. .184
     mat_peptide     185. .4699
                     /product="mature IgA protease (AA 1-1505)"
     misc_feature    3061. .3062
                     /note="autoproteolytic site for IgA protease"
     misc_feature    3157. .3158
                     /note="autoproteolytic site for IgA protease"
     misc_feature    3466. .3467
                     /note="autoproteolytic site for IgA protease"
     repeat_unit     4725. .4739
                     /note="inverted repeat A"
     terminator      4725. .4762
                     /note="put. transcription termination signal"
     repeat_unit     4747. .4762
                     /note="inverted repeat A'"
BASE COUNT     1617 a   1104 c   1080 g   1098 t
ORIGIN


  Query Match             4.3%;  Score 188.8;  DB 1;  Length 4899;
  Best Local Similarity   61.0%;  Pred. No. 3.2e-33;

  Matches  345;  Conservative    0;  Mismatches  212;  Indels    9;  Gaps    2;

Qy   1167 gaataatggagaaaatatttcctttattgacaaaggaaaaggtgaattgatacttaccag 1226
          ||| || ||| ||||| || |||||   ||||| || |       ||| || | |
Db   1192 GAACAACGGACAAAATGTTACCTTTGAAGACAACGGCA------CTTTGGTATTGAATCA 1245

Qy   1227 caacatcaaccaaggcgcgggcggtttgtattttgagggtaattttacggtctcgcctaa 1286
          ||||||||||||||||||||||  ||| |||  | ||  |||  || |||
Db   1246 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 1305

Qy   1287 aaacaacgaa---acgtggcaaggcgcgggcgttcatatcagtgatggcagtaccgttac 1343
          |||  ||  ||    ||  |||  |||  |||||  ||  || |    ||  ||||  |  ||
Db   1306 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 1365

Qy   1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgct 1403
          ||||  ||||  ||      |     ||||  |  ||  | ||||||||||||||||||  ||
Db   1366 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 1425

Qy   1404 ggttcaagccaaaggggaaaaccaaggctcggtcagcgtgggcgacggtaaagtcatctt 1463
            |  | | || ||| |  ||||||||      | |  || ||||||||||  || ||  |
Db   1426 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 1485

Qy   1464 agatcagcaggcggacgatcaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523
          ||||  | || || |    |    |||   || || ||     || ||||| | |||||
Db   1486 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 1545

Qy   1524 cggcaggggacggtgcaactgaatgccgataatcagttcaaccccgacaaactctattt 1583
          |||  | || ||  ||  | | |||      ||||||  | || |||||  || || |||||
Db   1546 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 1605

Qy   1584 cggctttcgcggcggacgtttggatttgaacgggcattcgctttcgttccaccgcattca 1643
          |||  || || ||||| |||||||||   || ||  ||     |  | ||  | | ||| |
Db   1606 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 1665

Qy   1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703
            ||    |||||||| |||   ||||||||||||||   |||| | |  |||||  | ||
Db   1666 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 1725

Qy   1704 cattacaggcaataaagatattacta 1729
          | || || ||       ||||| |
Db   1726 ACTAACGGGTAAATCTTTAATTACCA 1751


RESULT  30
A02796
LOCUS       A02796                  4899 bp    DNA     linear   PAT 10-MAY-1993
DEFINITION  N.gonorrhoeae gene for IgA-protease precursor.
ACCESSION   A02796
VERSION     A02796.1  GI:345152
KEYWORDS    IgA protease.
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   Meyer,T.F., Halter,R. and Pohlner,J.
  TITLE     Process for recovering proteins by means of gene technology using
            gram-negative host cells
  JOURNAL   Patent: EP 0254090-A 1 27-JAN-1988;
            Max-Planck-Gesellschaft zur Foerderung der Wissenschaften
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
     CDS             104. .4702
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA00270.1"
                     /db_xref="GI:345153"
                     /translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
```

KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQNRFLLTGGANLNGKI
SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
BASE COUNT 1617 a 1105 c 1079 g 1098 t
ORIGIN

Query Match 4.3%; Score 188.8; DB 6; Length 4899;
Best Local Similarity 61.0%; Pred. No. 3.2e-33;
Matches 345; Conservative 0; Mismatches 212; Indels 9; Gaps 2;

```
Qy   1167 gaataatggagaaaatatttcctttattgacaaaggaaaaggtgaattgatacttaccag 1226
          ||| || ||| ||||| || |||||   ||||| || |        ||| || | |
Db   1192 GAACAACGGACAAAATGTTACCTTTGAAGACAACGGCA------CTTTGGTATTGAATCA 1245

Qy   1227 caacatcaaccaaggcgcgggcggtttgtattttgagggtaattttacggtctcgcctaa 1286
          ||||||||||||||||||||||  ||| |||  | ||  |||  || |||
Db   1246 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 1305

Qy   1287 aaacaacgaa---acgtggcaaggcgcgggcgttcatatcagtgatggcagtaccgttac 1343
          ||| || ||    || ||| ||| |||||  || || |    || ||||  |  ||
Db   1306 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 1365

Qy   1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgct 1403
          |||| |||| ||     |    |||| |  || | ||||||||||||||||||  ||
Db   1366 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 1425

Qy   1404 ggttcaagccaaaggggaaaaccaaggctcggtcagcgtgggcgacggtaaagtcatctt 1463
            |  | | || ||| |  |||||||     | |  || ||||||||||  || ||  |
Db   1426 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 1485

Qy   1464 agatcagcaggcggacgatcaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523
            ||||  | || || |   |    |||   || || ||     || |||||  | |||||
Db   1486 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 1545

Qy   1524 cggcaggggacggtgcaactgaatgccgataatcagttcaaccccgacaaactctattt 1583
          |||  | || ||  ||  | | |||      ||||||  | || |||||  || || |||||
Db   1546 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 1605

Qy   1584 cggctttcgcggcggacgtttggatttgaacgggcattcgctttcgttccaccgcattca 1643
          ||| || || ||||| |||||||||    || ||  ||     |  | ||  | | ||| |
Db   1606 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 1665

Qy   1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703
          ||    |||||||| |||   ||||||||||||||   |||| | |  |||||  | ||
Db   1666 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 1725

Qy   1704 cattacaggcaataaagatattacta 1729
            | || || ||       |||||
Db   1726 ACTAACGGGTAAATCTTTAATTACCA 1751
```

RESULT 31
A12416
LOCUS A12416 4899 bp DNA linear PAT 05-OCT-1994
DEFINITION IgA-protease gene precursor.
ACCESSION A12416
VERSION A12416.1 GI:640585
KEYWORDS .
SOURCE Neisseria gonorrhoeae.
ORGANISM *Neisseria gonorrhoeae*
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 4899)
AUTHORS .
TITLE *PROCESS FOR THE GENETIC PRODUCTION OF PROTEINS BY USING OF* GRAM-NEGATIVE HOST CELLS
JOURNAL Patent: WO 8800238-A 12 14-JAN-1988;
FEATURES Location/Qualifiers
source 1. .4899
/organism="Neisseria *gonorrhoeae*"
/db_xref="taxon:485"
gene 104. .4702
/gene="IgA protease precursor"
CDS 104. .4702
/gene="IgA protease precursor"
/codon_start=1
/*transl_table*=11
/protein_id="CAA01030.1"
/db_xref="GI:640586"
/translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQNRFLLTGGANLNGKI
SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
BASE COUNT 1617 a 1105 c 1079 g 1098 t
ORIGIN

Query Match 4.3%; Score 188.8; DB 6; Length 4899;
Best Local Similarity 61.0%; Pred. No. 3.2e-33;
Matches 345; Conservative 0; Mismatches 212; Indels 9; Gaps 2;

```
Qy   1167 gaataatggagaaaatatttcctttattgacaaaggaaaaggtgaattgatacttaccag 1226
          ||| || ||| ||||| || |||||   ||||| || |        ||| || | |
Db   1192 GAACAACGGACAAAATGTTACCTTTGAAGACAACGGCA------CTTTGGTATTGAATCA 1245

Qy   1227 caacatcaaccaaggcgcgggcggtttgtattttgagggtaattttacggtctcgcctaa 1286
          ||||||||||||||||||||||  ||| |||  | ||  |||  || |||
Db   1246 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 1305

Qy   1287 aaacaacgaa---acgtggcaaggcgcgggcgttcatatcagtgatggcagtaccgttac 1343
```

```
Db   1306 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 1365

Qy   1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgct 1403

Db   1366 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 1425

Qy   1404 ggttcaagccaaaggggaaaaccaaggctcggtcagcgtgggcgacggtaaagtcatctt 1463

Db   1426 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 1485

Qy   1464 agatcagcaggcggacgatcaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523

Db   1486 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 1545

Qy   1524 cggcaggggacggtgcaactgaatgccgataatcagttcaaccccgacaaactctattt 1583

Db   1546 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 1605

Qy   1584 cggctttcgcggcggacgtttggatttgaacgggcattcgctttcgttccaccgcattca 1643

Db   1606 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 1665

Qy   1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703

Db   1666 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 1725

Qy   1704 cattacaggcaataaagatattacta 1729

Db   1726 ACTAACGGGTAAATCTTTAATTACCA 1751


RESULT 32
AR079202
LOCUS       AR079202                4899 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 5 from patent US 5965424.
ACCESSION   AR079202
VERSION     AR079202.1  GI:10005948
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   Ambrosius,D., Dony,C. and Rudolph,R.
  TITLE     Methods for making neisseria or hemophilus IgA protease and DNA
            encoding the proteases
  JOURNAL   Patent: US 5965424-A 5 12-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="unknown"
BASE COUNT     1617 a   1104 c   1080 g   1098 t
ORIGIN


  Query Match             4.3%;  Score 188.8;  DB 6;  Length 4899;
  Best Local Similarity   61.0%;  Pred. No. 3.2e-33;
  Matches  345;  Conservative    0;  Mismatches  212;  Indels    9;  Gaps    2;

Qy   1167 gaataatggagaaaatatttccttttattgacaaaggaaaaggtgaattgatacttaccag 1226

Db   1192 GAACAACGGACAAAATGTTACCTTTGAAGACAACGGCA------CTTTGGTATTGAATCA 1245

Qy   1227 caacatcaaccaaggcgcgggcggtttgtattttgagggtaattttacggtctcgcctaa 1286

Db   1246 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 1305

Qy   1287 aaacaacgaa---acgtggcaaggcgcgggcgttcatatcagtgatggcagtaccgttac 1343

Db   1306 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 1365

Qy   1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgct 1403

Db   1366 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 1425
```

```
Qy   1404 ggttcaagccaaaggggaaaaccaaggctcggtcagcgtgggcgacggtaaagtcatctt 1463

Db   1426 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 1485

Qy   1464 agatcagcaggcggacgatcaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523

Db   1486 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 1545

Qy   1524 cggcaggggacggtgcaactgaatgccgataatcagttcaaccccgacaaactctattt 1583

Db   1546 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 1605

Qy   1584 cggctttcgcggcggacgtttggatttgaacgggcattcgctttcgttccaccgcattca 1643

Db   1606 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 1665

Qy   1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703

Db   1666 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 1725

Qy   1704 cattacaggcaataaagatattacta 1729

Db   1726 ACTAACGGGTAAATCTTTAATTACCA 1751


RESULT 33
NGCOMLGN
LOCUS       NGCOMLGN                8393 bp    DNA     linear   BCT 18-NOV-1996
DEFINITION  N.gonorrhoeae gene for competence lipoprotein ComL.
ACCESSION   Z49895
VERSION     Z49895.1  GI:1107832
KEYWORDS    ComL; competence lipoprotein.
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 8393)
  AUTHORS   Fussenegger,M., Facius,D., Meier,J. and Meyer,T.F.
  TITLE     A novel peptidoglycan-linked lipoprotein (ComL) that functions in
            natural transformation competence of Neisseria gonorrhoeae
  JOURNAL   Mol. Microbiol. 19 (5), 1095-1105 (1996)
  MEDLINE   96249702
REFERENCE   2  (bases 1 to 8393)
  AUTHORS   Fussenegger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1995) Martin Fussenegger Dr., Infektionsbiologie,
            Max-Planck-Institut, fuer Biologie, Spemannstr. 34, Tuebingen, BW,
            D-72076, Germany
FEATURES             Location/Qualifiers
     source          1. .8393
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11A"
                     /db_xref="taxon:485"
     CDS             120. .923
                     /codon_start=1
                     /transl_table=11
                     /product="ComL, competence lipoprotein"
                     /protein_id="CAA90076.1"
                     /db_xref="GI:1107833"
                     /db_xref="SPTREMBL:Q50985"
                     /translation="MKKILLTVSLGLALSACATQGTADKDAQITQDWSVEKLYAEAQD
                     ELNSSNYTRAVKLYEILESRFPTSRHARQSQLDTAYAYYKDDEKDKALAAIERFRRLH
                     PQHPNMDYALYLRGLVLFNEDQSFLNKLASQDWSDRDPKANREAYQAFAELVQRFPNS
                     KYAADATARMVKLVDALGGNEMSVARYYMKRGAYIAAANRAKKIIGSYQNTRYVEESL
                     AILELAYKKLDKPQLAADTRRVLETNFPKSPFLTHAWQPDDMPWWRYWH"
     terminator      934. .967
BASE COUNT     2550 a   2300 c   1886 g   1657 t
ORIGIN


  Query Match             4.3%;  Score 188.8;  DB 1;  Length 8393;
  Best Local Similarity   61.0%;  Pred. No. 3.5e-33;
```

```
  Matches  345;  Conservative    0;  Mismatches  212;  Indels    9;  Gaps    2;

Qy   1167 gaataatggagaaaatatttccttttattgacaaaggaaaaggtgaattgatacttaccag 1226
          ||| || ||| ||||| || |||||    ||||| || |        ||| || | |
Db   4686 GAACAACGGACAAAATGTTACCTTTGAAGACAACGGCA------CTTTGGTATTGAATCA 4739

Qy   1227 caacatcaaccaaggcgcgggcggtttgtattttgagggtaattttacggtctcgcctaa 1286
           ||||||||||||||||||||||  ||| |||  | ||  |||  || |||
Db   4740 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 4799

Qy   1287 aaacaacgaa---acgtggcaaggcgcgggcgttcatatcagtgatggcagtaccgttac 1343
          ||| || ||    || |||  ||| |||||  || || |    || ||||  |  ||
Db   4800 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 4859

Qy   1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgct 1403
          |||| |||| ||      |     |||| |  || | ||||||||||||||||  ||
Db   4860 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 4919

Qy   1404 ggttcaagccaaaggggaaaaccaaggctcggtcagcgtgggcgacggtaaagtcatctt 1463
            | | | || ||| |  ||||||||     | |  || ||||||||||  || ||  |
Db   4920 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 4979

Qy   1464 agatcagcaggcggacgatcaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523
            ||||  | || || |   |    |||   || || ||     || ||||| | |||||
Db   4980 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 5039

Qy   1524 cggcaggggacggtgcaactgaatgccgataatcagttcaaccccgacaaactctattt 1583
          |||  | || ||  ||  | | |||      ||||||  | || |||||  || || |||||
Db   5040 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 5099

Qy   1584 cggctttcgcggcggacgtttggatttgaacgggcattcgctttcgttccaccgcattca 1643
          ||| || || ||||| |||||||||    || ||  ||     |  | ||  | | ||| |
Db   5100 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 5159

Qy   1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703
           ||    |||||||| |||   |||||||||||||   |||| | |  |||||  | ||
Db   5160 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 5219

Qy   1704 cattacaggcaataaagatattacta 1729
            | || || ||       ||||| |
Db   5220 ACTAACGGGTAAATCTTTAATTACCA 5245


RESULT  34
AF012206
LOCUS       AF012206                2992 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4024 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012206
VERSION     AF012206.1  GI:2529495
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2992
                     /organism="Neisseria meningitidis"
                     /strain="Z4024"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup VI; isolated in East-Germany,
                     1985"
     mat_peptide     1. .>2992
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2992
                     /gene="iga"
                     /allele="iga6"
     CDS             <1. .>2992
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45789.2"
                     /db_xref="GI:12597966"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     DANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQV
                     GIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTS
                     HASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGANLNGKISVTQGNVLLSGRPTPHARDFVNKSSAR
                     KDAHFSKNNEVVFEDDWINRTFKAAEITVNQSASFSSGRNVSDITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFGATQINGNVNLSQNAALVLGKAALW
                     GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKINH
                     LSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQDR
                     SRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQKA
                     TQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      983 a    684 c    639 g    686 t
ORIGIN


  Query Match             4.3%;  Score 188.4;  DB 1;  Length 2992;
  Best Local Similarity   58.4%;  Pred. No. 3.6e-33;
  Matches  349;  Conservative    0;  Mismatches  246;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          |||  || ||    ||| || |       ||||||||||||||||||||||||||| ||| ||
Db   1036 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1095

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     |  ||| || |     || |||| ||| |||||
Db   1096 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCGGGGA 1155

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1156 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1215

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1216 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1275

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  ||||  | || ||      |    |||   |
Db   1276 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTCC 1335

Qy   1496 aagccctttagtgaaatcggcttggtcagcggcagggggacggtgcaactgaatgccgata 1555
          | || ||      || ||||| | ||||||||  | || ||  ||  | | |||       |
Db   1336 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1395

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| ||  | ||||||||  || || ||||| |||||||||   || |
Db   1396 ATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1455

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||    ||||||||||
Db   1456 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1515
```

```
Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   |  | | |  |||||  | ||  | || |||||    |   |||||  |  |
Db   1516 ACAACACAAGCCACGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1575

Qy   1736 gcaataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcga 1793
            |||        |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1576 AAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTA 1633


RESULT  35
AF012210
LOCUS       AF012210               2995 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4099 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012210
VERSION     AF012210.1  GI:2529503
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Neisseria meningitidis"
                     /strain="Z4099"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup VIII; isolated in China,
                     1975"
     mat_peptide     <1. .>2995
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2995
                     /gene="iga"
                     /allele="iga8"
     CDS             <1. .>2995
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45793.2"
                     /db_xref="GI:12597970"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQHDNAGTIKGYREHHWNITSGTNSKIGSTAVRLAGNE
                     RNANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGNYTVKGANNDITWLGAGIDVADGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     DHTSTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     QKDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFDATQIHGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQERLNLFDASSVRD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRQVKPTHSPATNTASQAQK
                     ATQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      996 a    690 c    632 g    677 t

ORIGIN


  Query Match             4.3%;  Score 188.4;  DB 1;  Length 2995;
  Best Local Similarity   60.9%;  Pred. No. 3.6e-33;
  Matches  325;  Conservative    0;  Mismatches  206;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          |||  ||  ||    |||  ||  |       ||  ||||||||||||||||||||| ||| ||
Db   1039 AAGACAACGGCACTTTGGTATTGGATCAAAATATCAACCAAGGCGCGGGCGGTCTGTTTT 1098

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  ||||  ||  |||     |  ||| || ||    || |||| |||||||||
Db   1099 TCAAAGGCAATTACACAGTCAAAGGTGCAAATAATGACATCACCTGGCTAGGCGCGGGGA 1158

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1159 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1218

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1219 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1278

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  ||||  | || ||     |    |||  |
Db   1279 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTCC 1338

Qy   1496 aagcctttagtgaaaatcggcttggtcagcggcaaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||   || |||      |
Db   1339 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTGCTAAATAGTTCAA 1398

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db   1399 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1458

Qy   1616 ggcattcgctttcgttccaccgcattcaaaatacCgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1459 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1518

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattacta 1729
          |||||  |||| |    |||||  | ||  | || || ||       ||||| |
Db   1519 ACAATACAGACCACACCTCCACAATCACGCTAACGGGTAAATCTTTAATTACCA 1572


RESULT  36
NMIGA139
LOCUS       NMIGA139               1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (81139).
ACCESSION   X82477
VERSION     X82477.1  GI:732856
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="81139"
                     /db_xref="taxon:487"
```

```
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57860.1"
                     /db_xref="GI:732857"
                     /db_xref="SPTREMBL:Q51163"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHH
                     WKTTGTNSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYI"
BASE COUNT      498 a    320 c    321 g    350 t
ORIGIN
```

```
  Query Match             4.3%;  Score 188;  DB 1;  Length 1489;
  Best Local Similarity   60.9%;  Pred. No. 3.9e-33;
  Matches  324;  Conservative    0;  Mismatches  205;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db    926 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 985

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| |||||
Db    986 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1045

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1046 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1105

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1106 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1165

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db   1166 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1225

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||  | | |||      |
Db   1226 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1285

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db   1286 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1345

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| ||||  ||||||||||
Db   1346 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1405

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattac 1727
          ||||   || | | |  ||||| | ||  | || || ||       |||||
Db   1406 ACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGATTAC 1457
```

```
RESULT  37
NMIGA341
LOCUS       NMIGA341                1489 bp    DNA     linear   BCT 12-JUL-1999
DEFINITION  N.meningitidis iga gene (IHN5341).
ACCESSION   X82478
VERSION     X82478.1  GI:732862
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="IHN5341"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /pseudo
                     /transl_table=11
                     /codon_start=1
                     /product="IgA1 protease"
                     /db_xref="PID:e123968"
BASE COUNT      498 a    319 c    321 g    351 t
ORIGIN
```

```
  Query Match             4.2%;  Score 186.4;  DB 1;  Length 1489;
  Best Local Similarity   60.7%;  Pred. No. 9.4e-33;
  Matches  323;  Conservative    0;  Mismatches  206;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db    926 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 985

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| |||||
Db    986 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1045

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1046 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1105

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1106 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1165

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db   1166 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCT 1225

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||  | | |||      |
Db   1226 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1285

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db   1286 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1345

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1346 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1405

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattac 1727
          ||||   || | | |  ||||| | ||  | || || ||       |||||
Db   1406 ACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGATTAC 1457
```

```
RESULT  38
NMIGAG80
LOCUS       NMIGAG80                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NCG80).
ACCESSION   X82479
VERSION     X82479.1  GI:732883
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="NGC80"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57862.1"
                     /db_xref="GI:732884"
                     /db_xref="SPTREMBL:Q51172"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHH
                     WKTTGTNSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRRTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYI"
BASE COUNT      498 a    321 c    320 g    350 t
ORIGIN
```

```
  Query Match             4.2%;  Score 186.4;  DB 1;  Length 1489;
  Best Local Similarity   60.7%;  Pred. No. 9.4e-33;
  Matches  323;  Conservative    0;  Mismatches  206;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||| ||| ||
Db    926 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 985

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| |||||
Db    986 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1045

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1046 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1105

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1106 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1165

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || ||  |    |||  |
Db   1166 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1225

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||      || ||||| | |||||||||  |  | ||  ||  | | |||      |
Db   1226 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCCGTACATTGGTATTAAATAGTTCAA 1285

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db   1286 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1345

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | || ||| |  ||    ||||||||| |||   ||||||||||
Db   1346 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1405

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattac 1727
          ||||   || | | |  ||||  | ||  | || || ||       |||||
Db   1406 ACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGATTAC 1457
```

```
RESULT  39
AF012203
LOCUS       AF012203                2980 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z2491 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012203
VERSION     AF012203.1  GI:2529489
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2980)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2980)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2980
                     /organism="Neisseria meningitidis"
                     /strain="Z2491"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup IV-1; isolated in Gambia,
                     1983"
     mat_peptide     1. .>2980
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2980
                     /gene="iga"
                     /allele="iga4"
     CDS             <1. .>2980
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45786.1"
                     /db_xref="GI:2529490"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     GANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
```

VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
GIVSGRGTLVLNSPDQINPNNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
LKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAQ
KDAHFSKNNEVVFEDDWINRTFKATEITVNQSASFSSGRNVSNITANITATDNAKVNL
GYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATKINGNVNLNQNAALVLGKAALW
GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKINH
LSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVRDR
SHLSVSLANHHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQTD
SAQIAKPQNIVVAPPSPQAN"
BASE COUNT 974 a 687 c 642 g 677 t
ORIGIN

Query Match 4.2%; Score 185.2; DB 1; Length 2980;
Best Local Similarity 58.0%; Pred. No. 2e-32;
Matches 347; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

```
Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||| ||| ||
Db   1036 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1095

Qy   1259 ttgagggtaatttttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     |  ||| || |     || |||| ||| || ||
Db   1096 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCAGGGA 1155

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1156 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1215

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1216 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1275

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          |  |  || ||||||||||  || ||  |  |||| || || || |   |   |||   |
Db   1276 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1335

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | || ||||  | || ||  ||   ||||
Db   1336 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGGCGCGGTACATTGGTGCTGAACAGTCCGG 1395

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          | ||| | |||||  | || || |||||||| || || ||||| |||||||||   || |
Db   1396 ACCAGATTAACCCGAATAATCTGTATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1455

Qy   1616 ggcattcgctttcgttccaccgcattcaaaatacgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1456 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1515

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||  ||||   |  ||||| } ||  | || ||||||   |   |||||  |  |
Db   1516 ACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1575

Qy   1736 gcaataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcga 1793
          |||         |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1576 AAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTA 1633
```

RESULT 40
AF012204
LOCUS AF012204 2980 bp DNA linear BCT 30-JAN-2001
DEFINITION Neisseria meningitidis strain Z3906 IgA1 protease precursor (iga) gene, partial cds.
ACCESSION AF012204
VERSION AF012204.1 GI:2529491
KEYWORDS .
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 2980)
AUTHORS Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del Valle,J. and Achtman,M.
TITLE Clonal descent and *microevolution* of Neisseria meningitidis during 30 years of epidemic spread
JOURNAL Mol. Microbiol. 25 (6), 1047-1064 (1997)
MEDLINE 98010345
REFERENCE 2 (bases 1 to 2980)
AUTHORS Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del Valle,J. and Achtman,M.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES Location/Qualifiers
source 1. .2980
/organism="Neisseria meningitidis"
/strain="Z3906"
/db_xref="taxon:487"
/note="serogroup A, subgroup III; isolated *in China*, 1966; also present in strain B1, serogroup A, subgroup IV-2 isolated *in* USA, 1937"
mat_peptide 1. .>2980
/gene="iga"
/product="IgA1 protease"
gene <1. .>2980
/gene="iga"
/allele="iga2"
CDS <1. .>2980
/gene="iga"
/codon_start=3
/transl_table=11
/product="IgA1 protease precursor"
/protein_id="AAC45787.1"
/db_xref="GI:2529492"
/translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
GANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
GIVSGRGTLVLNSPDQINPNNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
LKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAQ
KDAHFSKNNEVVFEDDWINRTFKATEITVNQSASFSSGRNVSNITANITATDNAKVNL
GYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAALW
GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKINH
LSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVRDR
SHLSVSLANHHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQTD
SAQIAKPQNIVVAPPSPQAN"
BASE COUNT 973 a 688 c 642 g 677 t
ORIGIN

Query Match 4.2%; Score 185.2; DB 1; Length 2980;
Best Local Similarity 58.0%; Pred. No. 2e-32;
Matches 347; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

```
Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||| ||| ||
Db   1036 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1095

Qy   1259 ttgagggtaatttttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     |  ||| || |     || |||| ||| || ||
Db   1096 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCAGGGA 1155

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1156 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1215

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
```

```
Db   1216 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1275

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |    |    |||   |
Db   1276 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1335

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | || ||||| | || ||  ||   |||||
Db   1336 AGGCTTTCTCCCAAGTCGGCATTGTGAGCGGGCGCGGTACATTGGTGCTGAACAGTCCGG 1395

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          | ||| | |||||  | || || |||||||| || || ||||| |||||||||   || |
Db   1396 ACCAGATTAACCCGAATAATCTGTATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1455

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||     |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1456 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1515

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   ||||   |  |||||  | ||  | || |||||   |   |||||  |  |
Db   1516 ACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1575

Qy   1736 gcaataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcga 1793
            |||         |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1576 AAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTA 1633


RESULT  41
AF012205
LOCUS       AF012205                2980 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z3910 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012205
VERSION     AF012205.1  GI:2529493
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2980)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2980)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2980
                     /organism="Neisseria meningitidis"
                     /strain="Z3910"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup III; isolated in China, 1966"
     mat_peptide     1. .>2980
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2980
                     /gene="iga"
                     /allele="iga1"
     CDS             <1. .>2980
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45788.1"
                     /db_xref="GI:2529494"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSLEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     GANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
                     GIVSGRGTLVLNSPDQINPNNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
                     RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAQ
                     KDAHFSKNNEVVFEDDWINRTFKATEITVNQSASFSSGRNVSNITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAALW
                     GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKINH
                     LSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVRDR
                     SHLSVSLANHHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQTD
                     SAQIAKPQNIVVAPPSPQAN"
BASE COUNT      973 a    687 c    642 g    678 t
ORIGIN


  Query Match             4.2%;  Score 185.2;  DB 1;  Length 2980;
  Best Local Similarity   58.0%;  Pred. No. 2e-32;
  Matches  347;  Conservative    0;  Mismatches  248;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||| ||| ||
Db   1036 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1095

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     |  ||| || |     || |||| ||| || ||
Db   1096 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCAGGGA 1155

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||     |     |||| |
Db   1156 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1215

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1216 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1275

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  |||| || || || |    |    |||   |
Db   1276 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 1335

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | || ||||| | || ||  ||   |||||
Db   1336 AGGCTTTCTCCCAAGTCGGCATTGTGAGCGGGCGCGGTACATTGGTGCTGAACAGTCCGG 1395

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          | ||| | |||||  | || || |||||||| || || ||||| |||||||||   || |
Db   1396 ACCAGATTAACCCGAATAATCTGTATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1455

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||     |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1456 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1515

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   ||||   |  |||||  | ||  | || |||||   |   |||||  |  |
Db   1516 ACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1575

Qy   1736 gcaataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcga 1793
            |||         |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1576 AAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTA 1633


RESULT  42
NMA3Z2491/c
LOCUS       NMA3Z2491               311321 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 3/7.
ACCESSION   AL162754 AL157959
VERSION     AL162754.2  GI:7379424
KEYWORDS    .
```

```
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 311321)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 311321)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .311321
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     RBS             115. .120
     gene            129. .587
                     /gene="NMA0698"
     CDS             129. .587
                     /gene="NMA0698"
                     /note="NMA0698, possible periplasmic protein, len: 152 aa;
                     unknown, contains a probable N-terminal signal sequence
                     and lies within a region of unusually low GC content"
                     /codon_start=1
                     /transl_table=11
                     /product="putative periplasmic protein"
                     /protein_id="CAB83985.1"
                     /db_xref="GI:7379425"
                     /db_xref="SPTREMBL:Q9JQQ2"
                     /translation="MLTKLKILLFLFLFVFVLAINLLFFFSSDIESFGNYQFEYVYD
                     KGWPANYILVMKDGNEGNFDKIISGLVLEYYKEDDNIYFSYIDGQGFASDSCYYKPEI
                     LYGKIILNKNHIININSMEKNNFLSEDKIMKGTRNWLADPKNKCNIQTLD"
     gene            742. .1077
                     /gene="NMA0699"
     CDS             742. .1077
                     /gene="NMA0699"
                     /note="NMA0699, pseudogene, probable ABC transporter
                     protein, len: 336 bp; similar to C-termini of many ABC
                     transporters e.g. TR:Q46973 (EMBL:U47048), mtfB,
                     Escherichia coli microcin transport protein (707 aa),
                     fasta scores; E(): 2.6e-19, 53.2% identity in 111 aa
                     overlap. Contains PS00211 ABC transporters family
                     signature. NMA0686 may be the remainder of this
                     pseudogene"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="putative ABC transporter protein (pseudogene)"
     misc_feature    868. .912
                     /gene="NMA0699"
                     /note="PS00211 ABC transporters family signature"
     gene            complement(1101. .2327)
                     /gene="NMA0700"
     CDS             complement(1101. .2327)
                     /gene="NMA0700"
                     /note="NMA0700, possible ribonuclease BN, len: 408 aa;
                     similar to SW:RBN_ECOLI (EMBL:L19201), rbn, Escherichia
                     coli ribonuclease BN (EC 3.1.-.-), fasta scores; E():
                     2.3e-28, 34.3% identity in 268 aa overlap. Longer than rbn
                     at the C-terminus. Also similar to TR:O85449
                     (EMBL:AF067083) Vitreoscilla sp. hypothetical protein (376
                     aa), fasta scores; E(): 0, 45.0% identity in 333 aa
                     overlap. Contains a region similar to NMA0677, fasta
                     scores; E(): 4.3e-08, 64.4% identity in 45 aa overlap.
                     Contains hydrophobic, possible membrane-spanning regions,
                     as does rbn"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribonuclease BN"
                     /protein_id="CAB83987.1"
                     /db_xref="GI:7379426"
                     /db_xref="SPTREMBL:Q9JVT9"
                     /translation="MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLL
                     ALVPVLTVMVAVASIFPVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTA
                     IGSVMLVVTSLMLIRTIDNTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFM
                     VGSVQDAALASGAPQWSGALRTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATA
                     FCLETARSLFTWYMGNFDGYRSIYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQ
                     GEAFRRGFDSRGRFDDVLKILLLLDAAQKEGKALPVQEFRRHINMGYDELGELLEKLA
                     RHGYIYSGRQGWVLKTGADSIELNELFKLFVYRPLPVERDHVNQAVDAVMMPCLQTLN
                     MTLAEFDAQAKKQQQS"
     misc_feature    1376. .1385
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(2417. .2426)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2444. .2453
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            2486. .2743
                     /gene="NMA0701"
     CDS             2486. .2743
                     /gene="NMA0701"
                     /note="NMA0701, possible pseudogene, len: 258 bp; shows
                     weak similarity to part of SW:WRBA_ECOLI (EMBL:M99166),
                     wrbA, Escherichia coli Trp repressor binding protein (197
                     aa), fasta scores; E(): 0.064, 29.3% identity in 82 aa
                     overlap. Also similar to part of TR:O85450 (EMBL:AF067083)
                     Vitreoscilla sp. Trp repressor binding protein (fragment)
                     (124 aa), fasta scores; E(): 6.6e-14, 56.6% identity in 83
                     aa overlap"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="hypothetical protein NMA0701 (pseudogene)"
     misc_feature    complement(2755. .2764)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2778. .2787
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             2861. .2864
     gene            2871. .3530
                     /gene="NMA0702"
     CDS             2871. .3530
                     /gene="NMA0702"
                     /note="NMA0702, len: 219 aa; similar to TR:Q52925
                     (EMBL:Z50189), exsB, Rhizobium meliloti putative regulator
                     of succinoglycan biosynthesis (not a transcriptional
                     regulator) (234 aa), fasta scores; E(): 3.1e-15, 32.7%
                     identity in 205 aa overlap. Also similar to many
                     hbacterial ypothetical proteins e.g. SW:YBAX_HAEIN
                     (EMBL:U32798), HI1191, Haemophilus influenzae hypothetical
                     protein (196 aa), fasta scores; E(): 0, 79.2% identity in
                     173 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0702"
                     /protein_id="CAB83989.1"
                     /db_xref="GI:7379427"
```

```
                     /db_xref="SPTREMBL:Q9JVT8"
                     /translation="MSNQKALVIFSGGQDSTTCLIQAIQTYGRENVQAITFQYGQRHA
                     VELERARWIAQDLGVKQTVLDLSLMRQITHNALMDDTAAIETAENGVPNTFVDGRNAL
                     FLLYAAIYAKGQGIRHIIAGVCETDFSGYPDCRDVFVKSMNVTLNLAMDYDFQIHTPL
                     MYLTKAQTWALADEMGALDYIREQTHTCYNGIVGGCRECPSCILRERGLAEYLESKKA
                     V"
     misc_feature    3522. .3531
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             3547. .3550
     gene            3558. .4076
                     /gene="NMA0703"
     CDS             3558. .4076
                     /gene="NMA0703"
                     /note="NMA0703, len: 172 aa; unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0703"
                     /protein_id="CAB83990.1"
                     /db_xref="GI:7379428"
                     /db_xref="SPTREMBL:Q9JVT7"
                     /translation="MPKLHMFYLGGNAGRSNIEVHDIQFAVCDDYREAVPALKAAWFG
                     DADKIHIDGWQVVEWADGYDIAVSETPKTKMPSENAPRLYFANVGGYRAGQLAEAHAF
                     GLFAAATPAEAKQKALQTLLTDSYVQQHKDNLKDVDNLLALDRIGNFHIRLTPNPHGK
                     PAEIGFQGYLPI"
     misc_feature    3782. .3791
                     /gene="NMA0703"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            4083. .4505
                     /gene="NMA0704"
     misc_feature    4083. .4502
                     /gene="NMA0704"
                     /note="Pfam match to entry PF01242 PTPS, 6-pyruvoyl
                     tetrahydropterin synthase, score 10.80, E-value 1.7e-06"
     CDS             4083. .4505
                     /gene="NMA0704"
                     /note="NMA0704, len: 140 aa; similar to many hypothetical
                     proteins e.g. SW:YB90_HAEIN (EMBL:U32798), HI1190,
                     Haemophilus influenzae hypothetical protein (141 aa),
                     fasta scores; E(): 5.1e-32, 58.0% identity in 138 aa
                     overlap. Shows very weak similarity to eukaryotic
                     6-pyruvoyl-tetrahydropterin synthases e.g. SW:PTPS_RAT
                     (EMBL:M77850), pts, Rattus norvegicus 6-pyruvoyl
                     tetrahydrobiopterin synthase precursor (EC 4.6.1.10) (144
                     aa), fasta scores; E(): 0.0062, 27.3% identity in 154 aa
                     overlap. Contains Pfam match to entry PF01242 PTPS,
                     6-pyruvoyl tetrahydropterin synthase"

  Query Match             4.2%;  Score 185.2;  DB 1;  Length 311321;
  Best Local Similarity   58.0%;  Pred. No. 4.6e-32;
  Matches  347;  Conservative    0;  Mismatches  248;  Indels    3;  Gaps    1;

Qy    1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
           ||   || ||    ||| || |       ||||||||||||||||||||||| ||| ||
Db  188899 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 188840

Qy    1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
           |  | ||  |||  || |||     |  ||| || |     || |||| ||| || ||
Db  188839 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCAGGGA 188780

Qy    1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
           || || |    || ||||  |  ||   |||| |||| ||     |     |||| |
Db  188779 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 188720

Qy    1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
           || | ||||||||||||||||  |    |  | | || ||| |  ||||||||
Db  188719 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 188660

Qy    1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
           | |  || ||||||||||  || ||  |  |||| || || || |   |    |||   |
Db  188659 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTCC 188600

Qy    1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
           | || ||     || ||||| | || |||||  | || ||  ||   |||||
Db  188599 AGGCTTTCTCCCAAGTCGGCATTGTGAGCGGGCGCGGTACATTGGTGCTGAACAGTCCGG 188540

Qy    1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
           | ||| | |||||  | || || |||||||| || || ||||| |||||||||   || |
Db  188539 ACCAGATTAACCCGAATAATCTGTATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 188480

Qy    1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
           |  ||    |  | ||  | | ||| |  ||    ||||||||| |||   ||||||||||
Db  188479 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 188420

Qy    1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
           ||||   ||||    |  ||||| | ||  | || |||||   |   |||||  |  |
Db  188419 ACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 188360

Qy    1736 gcaataacaacaacttggatagcaaaaaagaaattgcctacaacggttggtttggcga 1793
             |||         |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db  188359 AAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTA 188302


RESULT  43
NMIGAC65
LOCUS       NMIGAC65                1525 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NGC65).
ACCESSION   X82484
VERSION     X82484.1  GI:732871
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1525)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1525)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1525
                     /organism="Neisseria meningitidis"
                     /strain="NGC65"
                     /db_xref="taxon:487"
     gene            1. .1525
                     /gene="iga"
     CDS             <1. .>1525
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57867.1"
                     /db_xref="GI:732872"
                     /db_xref="SPTREMBL:Q51168"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYA
                     VSVKHAKAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWSASNLGRLEDYNM
                     ARFNKFVTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKG
                     YDLRDLSQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALT
                     NYGVLGDSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDN
                     AGTIKGNGEHHWNITFGTNSHIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQ
                     GAGGLFFKGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEIN
                     GTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQVGIVSGRRTLVLNSSNQINPDNLYF
                     GFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITAPQNLSVYEI
                     "
BASE COUNT      510 a    320 c    330 g    365 t
ORIGIN
```

```
  Query Match             4.2%;  Score 185;  DB 1;  Length 1525;
  Best Local Similarity   60.0%;  Pred. No. 2e-32;
  Matches  327;  Conservative    0;  Mismatches  215;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          |||  || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db    962 AAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 1021

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||      |  ||| || ||    || |||| ||| |||||
Db   1022 TCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGGA 1081

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |      ||| |
Db   1082 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGTGACAGAT 1141

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |    |  | | || ||| |  ||||||||  :.
Db   1142 TGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1201

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || |||||||| | ||| ||  |  ||||  | || ||     |   |||   |
Db   1202 TAAAAGTCGGCGACGGCACAGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTCC 1261

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | |||||||||  |  | ||  ||  | | |||     |
Db   1262 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCCGTACATTGGTATTAAATAGTTCAA 1321

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db   1322 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1381

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1382 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1441

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattactacaaccg 1735
          ||||   || | | |  ||||| | ||  | || ||||    |   |||||  |  | |
Db   1442 ACAACACAGGCCACGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCTC 1501

Qy   1736 gcaat 1740
            |||
Db   1502 AAAAT 1506


RESULT  44
NMIGAF48
LOCUS       NMIGAF48               1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (HF48).
ACCESSION   X82480
VERSION     X82480.1  GI:732875
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="HF48"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57863.1"
                     /db_xref="GI:732876"
                     /db_xref="SPTREMBL:Q57035"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQKADSNQKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHTSTITLTGKSLITNPNSLSVHSI"
BASE COUNT      496 a    319 c    325 g    349 t
ORIGIN


  Query Match             4.2%;  Score 183.6;  DB 1;  Length 1489;
  Best Local Similarity   60.3%;  Pred. No. 4.3e-32;
  Matches  322;  Conservative    0;  Mismatches  209;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db    926 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 985

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||      || ||| || ||    || |||| ||| |||||
Db    986 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1045

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1046 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1105

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |      |  | | || ||| |  ||||||||
Db   1106 TGGCGAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1165

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || ||||||||||  || ||  |  ||||  | || ||     |   |||   |
Db   1166 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTCC 1225

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||      || ||||| | ||||||||  | || ||  ||  | | |||     |
Db   1226 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCCGGTACATTGGTATTAAATAGTTCAA 1285

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db   1286 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1345

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1346 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1405

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattacta 1729
          ||||   || | |    ||||  | ||  | || || ||       ||||| |
Db   1406 ACAACACAGGCCACACCTCCACAATCACGCTAACGGGTAAATCTTTAATTACCA 1459


RESULT  45
NMIGAF96
LOCUS       NMIGAF96               1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (HF96).
ACCESSION   X82475
VERSION     X82475.1  GI:732879
KEYWORDS    iga gene; IgA1 protease.
```

```
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="HF96"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57858.1"
                     /db_xref="GI:732880"
                     /db_xref="SPTREMBL:Q57035"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQKADSNQKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHTSTITLTGKSLITNPNSLSVHSI"
BASE COUNT      496 a    319 c    325 g    349 t
ORIGIN
```

```
  Query Match             4.2%;  Score 183.6;  DB 1;  Length 1489;
  Best Local Similarity   60.3%;  Pred. No. 4.3e-32;
  Matches  322;  Conservative    0;  Mismatches  209;  Indels    3;  Gaps    1;

Qy   1199 aaggaaaaggtgaattgatacttaccagcaacatcaaccaaggcgcgggcggtttgtatt 1258
          ||   || ||    ||| || |       ||||||||||||||||||||||||| ||| ||
Db    926 AAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTTT 985

Qy   1259 ttgagggtaattttacggtctcgcctaaaaacaacgaa---acgtggcaaggcgcgggcg 1315
          |  | ||  |||  || |||     || ||| || ||    || |||| ||| |||||
Db    986 TCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGGA 1045

Qy   1316 ttcatatcagtgatggcagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcc 1375
          || || |    || ||||  |  ||   |||| |||| ||      |     |||| |
Db   1046 TTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGAT 1105

Qy   1376 tgtccaaaatcggcaaaggcacgctgctggttcaagccaaaggggaaaaccaaggctcgg 1435
          || | ||||||||||||||||  |     |  | | || ||| |  ||||||||
Db   1106 TGGCGAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAAT 1165

Qy   1436 tcagcgtgggcgacggtaaagtcatcttagatcagcaggcggacgatcaaggcaaaaaac 1495
          | |  || |||||||||  || ||  |  ||||  | || ||     |    |||   |
Db   1166 TAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTCC 1225

Qy   1496 aagcctttagtgaaatcggcttggtcagcggcaggggacggtgcaactgaatgccgata 1555
          | || ||     || ||||| | ||||||||  | || ||  ||  | | |||      |
Db   1226 AGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCAA 1285

Qy   1556 atcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttgaacg 1615
          ||||| | || ||||| || || |||||||| || || ||||| |||||||||   || |
Db   1286 ATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAATG 1345

Qy   1616 ggcattcgctttcgttccaccgcattcaaaataccgatgaaggggcgatgattgtcaacc 1675
          |  ||    |  | ||  | | ||| |  ||    |||||||| |||   ||||||||||
Db   1346 GCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAACC 1405

Qy   1676 acaatcaagacaaagaatccaccgttaccattacaggcaataaagatattacta 1729
          ||||   || | |    |||||  | ||  | || || ||       ||||| |
Db   1406 ACAACACAGGCCACACCTCCACAATCACGCTAACGGGTAAATCTTTAATTACCA 1459


Search completed: June 30, 2002, 14:11:55
Job time: 75333 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 10:28:37 ; Search time 27321.5 Seconds
(without alignments)
3350.209 Million cell updates/sec

Title: US-09-303-518D-649
Perfect score: 4374
Sequence: 1 atgaaaacaaccgacaaacg.....aattaggctaccgctggtaa 4374

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4374	100.0	4374	6	A96616	A96616 Sequence 64
2	4374	100.0	4374	6	AX044007	AX044007 Sequence
c 3	4374	100.0	10074	1	AE002547	AE002547 Neisseria
c 4	4374	100.0	349980	6	AX044034	AX044034 Sequence
5	4366	99.8	4374	1	NME242535	AJ242535 Neisseria
6	4266	97.5	4386	1	NME296277	AJ296277 Neisseria
7	3954.8	90.4	4407	6	A96620	A96620 Sequence 65
8	3566.8	81.5	4350	1	NME296276	AJ296276 Neisseria
9	3566.8	81.5	349061	1	NMA2Z2491	AL162753 Neisseria
10	3508	80.2	4350	6	A96618	A96618 Sequence 65
11	2693.2	61.6	4355	6	A96614	A96614 Sequence 64
c 12	1274.4	29.1	13818	1	U32710	U32710 Haemophilus
13	1260.8	28.8	4317	1	HIU11024	U11024 Haemophilus
14	1249.4	28.6	4319	6	AR157320	AR157320 Sequence
15	318.8	7.3	578	1	NLIGA3327	AJ001740 Niesseria
16	313.6	7.2	584	1	NLIGA3354	AJ001739 Niesseria
17	280.6	6.4	572	1	NLIGA823	AJ001737 Niesseria
18	280.2	6.4	572	1	NLIGA3293	AJ001738 Niesseria
19	202.6	4.6	2992	1	AF012207	AF012207 Neisseria
20	197.8	4.5	2992	1	AF012206	AF012206 Neisseria
c 21	196.2	4.5	10485	1	AE002424	AE002424 Neisseria
c 22	196.2	4.5	349980	6	AX044030	AX044030 Sequence
23	195.6	4.5	1489	1	NMIGA771	X82468 N.meningiti
24	195.6	4.5	1489	1	NMIGATH2	X82469 N.meningiti
25	195.6	4.5	1525	1	NMIGAG117	X82483 N.meningiti
26	195.6	4.5	4899	1	NGIGAPRO	X04835 Neisseria g
27	195.6	4.5	4899	6	A02796	A02796 N.gonorrhoe
28	195.6	4.5	4899	6	A12416	A12416 IgA-proteas
29	195.6	4.5	4899	6	AR079202	AR079202 Sequence
30	195.6	4.5	8393	1	NGCOMLGN	Z49895 N.gonorrhoe
31	193	4.4	1492	1	NMIGA476	X82481 N.meningiti
32	193	4.4	2980	1	AF012203	AF012203 Neisseria
33	193	4.4	2980	1	AF012204	AF012204 Neisseria
34	193	4.4	2980	1	AF012205	AF012205 Neisseria
35	193	4.4	2983	1	AF012208	AF012208 Neisseria
36	193	4.4	2995	1	AF012209	AF012209 Neisseria
37	193	4.4	4659	1	AF235032	AF235032 Neisseria
c 38	193	4.4	311321	1	NMA3Z2491	AL162754 Neisseria
39	192.4	4.4	1525	1	NMIGAC65	X82484 N.meningiti
40	191.4	4.4	1489	1	NMIGA139	X82477 N.meningiti
41	191.4	4.4	1525	1	NMIGAC16	X82485 N.meningiti
42	191.4	4.4	2995	1	AF012210	AF012210 Neisseria
43	189.8	4.3	1489	1	NMIGA341	X82478 N.meningiti
44	189.8	4.3	1489	1	NMIGAG80	X82479 N.meningiti
45	188.2	4.3	1489	1	NMIGAF48	X82480 N.meningiti
46	188.2	4.3	1489	1	NMIGAF96	X82475 N.meningiti
47	188.2	4.3	2980	1	AF012211	AF012211 Neisseria
48	188.2	4.3	5261	1	NMIGAF13	X82474 N.meningiti
49	186.6	4.3	1489	1	NMIGA894	X82476 N.meningiti
50	186.6	4.3	1522	1	NMIGAO93	X82482 N.meningiti
51	186	4.3	1489	1	NMIGA027	X82472 N.meningiti
52	186	4.3	1489	1	NMIGA117	X82470 N.meningiti
53	186	4.3	1489	1	NMIGA159	X82471 N.meningiti
54	184.4	4.2	1489	1	NMIGAF54	X82473 N.meningiti
55	183.4	4.2	1525	1	NMIGA166	X82486 N.meningiti
56	175.2	4.0	6021	1	HEAIGA1PTC	M87491 Haemophilus
57	174.4	4.0	11288	1	AE002549	AE002549 Neisseria
58	174.4	4.0	172325	6	AX044035	AX044035 Sequence
59	174.4	4.0	349980	6	AX044034	AX044034 Sequence
60	166.2	3.8	5536	1	HEAIGA1PTA	M87489 Haemophilus
61	166.2	3.8	5657	1	HIIGA1P	X59800 Haemophilus
c 62	166.2	3.8	15212	1	U32779	U32779 Haemophilus
63	161.8	3.7	5088	1	HEAIGA1PTB	M87490 Haemophilus
64	151.8	3.5	5719	1	AF369907	AF369907 Haemophil
65	149	3.4	5096	1	HEAIGA1PTD	M87492 Haemophilus
66	149	3.4	5096	1	HIGA	X64357 H.influenza
c 67	141.6	3.2	213494	1	SFPWR100	AL391753 Shigella
68	140	3.2	4484	1	SFSEPAGN	Z48219 Shigella fl
c 69	138.6	3.2	37710	1	ECO278144	AJ278144 Escherich
70	135	3.1	4113	6	AX276281	AX276281 Sequence
c 71	135	3.1	48254	6	AX276220	AX276220 Sequence
72	134	3.1	4394	1	SFU35656	U35656 Shigella fl
73	134	3.1	5639	1	AF097644	AF097644 Escherich

c 74	134	3.1	51334	1	AF200692	AF200692 Shigella
75	127	2.9	1390	1	HIIGA635	X82488 H.influenza
76	126	2.9	5208	1	AF422145	AF422145 Escherich
77	118.4	2.7	4128	6	AX275950	AX275950 Sequence
c 78	118.4	2.7	7654	6	AX275945	AX275945 Sequence
79	118.2	2.7	9689	7	AF151674	AF151674 Bacteriop
80	118.2	2.7	11279	7	AF151091	AF151091 Prophage
81	112.4	2.6	1413	1	HIIGA284	X82487 H.influenza
82	108.8	2.5	10587	1	AF218073	AF218073 Escherich
83	108.8	2.5	12606	1	ECAJ3631	AJ223631 Escherich
84	94.8	2.2	1380	1	HIDNAIGA	X86103 H.influenza
85	90.2	2.1	1477	1	HIIGAA66	X82467 H.influenza
c 86	75.8	1.7	1098	1	SF2AETX1	Z47381 S.flexneri
c 87	61	1.4	125020	9	AF429315	AF429315 Homo sapi
c 88	60.6	1.4	7218	6	I66494	I66494 Sequence 14
89	59.4	1.4	144	6	A12818	A12818 IgA-proteas
90	58.4	1.3	2070	5	BSU43200	U43200 Boreogadus
91	56.8	1.3	1316	14	AY034637	AY034637 Equine he
92	56.4	1.3	1226	14	AY034642	AY034642 Equine he
93	56.4	1.3	125020	9	AF429315	AF429315 Homo sapi
c 94	55.6	1.3	15776	1	AE007069	AE007069 Mycobacte
c 95	55.6	1.3	37586	1	MTCY270	Z95388 Mycobacteri
c 96	55.6	1.3	37586	6	AX191745	AX191745 Sequence
97	53	1.2	347050	1	AL591981	AL591981 Listeria
98	52.6	1.2	9553	5	AF134321	AF134321 Dissostic
99	52.4	1.2	313450	1	AL596170	AL596170 Listeria
c 100	51.6	1.2	18471	1	AE007047	AE007047 Mycobacte

ALIGNMENTS

```
RESULT   1
A96616
LOCUS       A96616                   4374 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 649 from Patent WO9924578.
ACCESSION   A96616
VERSION     A96616.1  GI:6780200
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4374)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 649 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     1266 a   1163 c   1066 g    879 t
ORIGIN
```

Query Match 100.0%; Score 4374; DB 6; Length 4374;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy     421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 480

Qy     481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 540

Qy     541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600

Qy     601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660

Qy     661 gcaagtgcgtattcttggctcgttggtggcaataccttgcacaaaatggatcaggtggt 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy     721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggttttttaccaaca 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840

Qy     841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

Qy     901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960

Qy     961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 1020

Qy    1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080

Qy    1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140

Qy    1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacgaa 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200

Qy    1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 1260

Qy    1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1261 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320

Qy   1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380

Qy   1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440

Qy   1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500

Qy   1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560

Qy   1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1620

Qy   1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680

Qy   1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggcaat 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 1740

Qy   1741 aacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800

Qy   1801 acgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcacc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860

Qy   1861 ctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactg 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920

Qy   1921 tttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 1980

Qy   1981 aaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacattt 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 2040

Qy   2041 aaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaa 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 2100

Qy   2101 gtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 2160

Qy   2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaa 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 2220

Qy   2221 accattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 2280

Qy   2281 gatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaatctt 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 2340

Qy   2341 agtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaacctt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 2400

Qy   2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460

Qy   2461 gcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacg 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520

Qy   2521 ctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctcccta 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580

Qy   2581 gccgataaggcagtattccattttgaaagcagccgctttaccggacaaatcagcggcggc 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 2640

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaatta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcg 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 2760

Qy   2761 gcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgc 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 2820

Qy   2821 cgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 2880

Qy   2881 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2940

Qy   2941 cgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaac 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000

Qy   3001 aataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaac 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 3060

Qy   3061 aaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcg 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 3120

Qy   3121 tggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3180

Qy   3181 caagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaagacaac 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC 3240

Qy   3241 gcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagacagaaagc 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3300

Qy   3301 gttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcggaggaa 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA 3360

Qy   3361 gagaaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaagcggaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3420

Qy   3421 acccggccggctaccaccgccttccccccgcgcccgccgcgcccgccgggatttgccgcaa 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 ACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3480
```

```
Qy   3481 ctgcaaccccaaccgcagccccaaccgcagcgcgacctgatcagccgttatgccaatagc 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGC 3540

Qy   3541 ggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacaggacgaattagac 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGAC 3600

Qy   3601 cgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3660

Qy   3661 cactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggt 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3720

Qy   3721 atgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaa 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAA 3780

Qy   3781 aacaccttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcggg 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGG 3840

Qy   3841 caatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggc 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 CAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGC 3900

Qy   3901 agcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcag 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 AGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3960

Qy   3961 gcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgc 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGC 4020

Qy   4021 tatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggccctt 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTT 4080

Qy   4081 gcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacac 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACAC 4140

Qy   4141 atttccatcacgccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccga 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 ATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGA 4200

Qy   4201 acacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgcagtgcggaatgg 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 ACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 4260

Qy   4261 ggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGC 4320

Qy   4321 ccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 CCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4374


RESULT   2
AX044007
LOCUS       AX044007                4374 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 86 from Patent WO0066791.
ACCESSION   AX044007
VERSION     AX044007.1  GI:11342893
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4374)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 86 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
BASE COUNT     1266 a   1163 c   1066 g    879 t
ORIGIN


  Query Match             100.0%;  Score 4374;  DB 6;  Length 4374;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4374;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy     421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 480

Qy     481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 540

Qy     541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600

Qy     601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660

Qy     661 gcaagtgcgtattcttggctcgttggtggcaatacctttgcacaaaatggatcaggtggt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy     721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggtttttaccaaca 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
```

```
Qy    841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

Qy    901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagataccc attcagta 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960

Qy    961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 1020

Qy   1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080

Qy   1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140

Qy   1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttccttttattgacgaa 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200

Qy   1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 1260

Qy   1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320

Qy   1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380

Qy   1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440

Qy   1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500

Qy   1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560

Qy   1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1620

Qy   1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680

Qy   1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggcaat 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 1740

Qy   1741 aacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800

Qy   1801 acgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcacc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860

Qy   1861 ctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactg 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920

Qy   1921 tttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 1980

Qy   1981 aaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacattt 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 2040

Qy   2041 aaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaa 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 2100

Qy   2101 gtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 2160

Qy   2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaa 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 2220

Qy   2221 accattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 2280

Qy   2281 gatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaatctt 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 2340

Qy   2341 agtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaacctt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 2400

Qy   2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460

Qy   2461 gcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacg 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520

Qy   2521 ctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctcccta 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580

Qy   2581 gccgataaggcagtattccattttgaaagcagccgctttaccggacaaatcagcggcggc 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 2640

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaatta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcg 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 2760

Qy   2761 gcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgc 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 2820

Qy   2821 cgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 2880

Qy   2881 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2940

Qy   2941 cgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaac 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000

Qy   3001 aataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaac 3060
```

```
Db   3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 3060

Qy   3061 aaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcg 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 3120

Qy   3121 tggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3180

Qy   3181 caagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaagacaac 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC 3240

Qy   3241 gcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagacagaaagc 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3300

Qy   3301 gttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcggaggaa 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA 3360

Qy   3361 gagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaagcggaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3420

Qy   3421 acccggccggctaccaccgccttcccccgcgcccgccgcgcccgccgggatttgccgcaa 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 ACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3480

Qy   3481 ctgcaaccccaaccgcagccccaaccgcagcgcgacctgatcagccgttatgccaatagc 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGC 3540

Qy   3541 ggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacaggacgaattagac 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGAC 3600

Qy   3601 cgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3660

Qy   3661 cactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggt 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3720

Qy   3721 atgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaa 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAA 3780

Qy   3781 aacaccttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcggg 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGG 3840

Qy   3841 caatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggc 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 CAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGC 3900

Qy   3901 agcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcag 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 AGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3960

Qy   3961 gcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgc 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGC 4020

Qy   4021 tatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggcctt 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTT 4080

Qy   4081 gcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacac 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACAC 4140

Qy   4141 atttccatcacgccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccga 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 ATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGA 4200

Qy   4201 acacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgcagtgcggaatgg 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 ACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 4260

Qy   4261 ggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGC 4320

Qy   4321 ccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 CCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4374
```

```
RESULT   3
AE002547/c
LOCUS       AE002547               10074 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 189 of 206
            of the complete genome.
ACCESSION   AE002547 AE002098
VERSION     AE002547.1  GI:7227244
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10074)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10074)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10074
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            704. .1375
                     /gene="NMB1983"
     CDS             704. .1375
                     /gene="NMB1983"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="hypothetical protein"
                     /protein_id="AAF42311.1"
                     /db_xref="GI:7227245"
                     /translation="MFPDKYKLSLKENIFLAKKVLVAQIHNLSRFENCQTTLLQTEQI
                     IHGKNVASASLEDIQTILNLKRAYQYVISHISNGEPVDISLLKKINNIVAKDDSLAPG
                     DFRTGSVGVTLLDGSRHAPNPVKEIEVARVLQNIGLQSGSTTEAAVRFMLYCMRQQVF
                     WDGNKRTATLFANGLMMAGGCGILEISEMQMPQFNEKLSAFYRSGDDTDISKFVYQNC
                     ISGID"
gene                 1437. .2548
                     /gene="NMB1984"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     SP:Q00840 GB:Z11857 PID:45066 percent identity: 97.40;
                     identified by sequence similarity; putative;IS1106
                     transposase, authentic frameshift"
gene                 complement(2864. .7237)
                     /gene="NMB1985"
CDS                  complement(2864. .7237)
                     /gene="NMB1985"
                     /note="similar to GP:1408569 percent identity: 76.08;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion and penetration protein"
                     /protein_id="AAF42312.1"
                     /db_xref="GI:7227246"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVFYEPRQNGKYSFND
                     DNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINQGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNG
                     NLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHDAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFR
                     FMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFT
                     LQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALI
                     AAGRDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATT
                     AFPRARRARRDLPQLQPQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFA
                     EDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENT
                     FDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVLHYGIQ
                     ARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA
                     QHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAA
                     AAKGPQLEAQHSAGIKLGYRW"
gene                 complement(7371. .7802)
                     /gene="NMB1986"
CDS                  complement(7371. .7802)
                     /gene="NMB1986"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF42313.1"
                     /db_xref="GI:7227247"
                     /translation="MTENAQDKARQAVETVVKSPELVEQILSDEYVQIMIARCFHSGP
                     LPPPSDLAQYNDIISNGADRIMAMAEKEQAVRHETIRQDQTFNRRGQLYGFISAILIL
                     LFAVFLVWSGYPATAASLAGGTVFALAGAFVIGRSRDQGKN"
gene                 8087. .9433
                     /gene="NMB1987"
CDS                  8087. .9433
                     /gene="NMB1987"
                     /note="similar to GB:L10328 SP:P25522 PID:290554 GB:U00096
                     PID:2367268 percent identity: 69.20; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="thiophene and furan oxidation protein ThdF"
                     /protein_id="AAF42314.1"
                     /db_xref="GI:7227248"
                     /translation="MSDNVPTIAAVATAPGRGGVGVIRISGKNLLPMAQALCGKTPKP
                     RTATYADFTDTDGQAIDSGLLLFFAAPASFTGEDVIELQGHGGPVVMDMLLNRCLELG
                     ARLAEPGEFTKRAFLNDKLDLAQAEGVADLIDASSRSAARLALRSLKGDFSRRIHGLV
                     EDLITLRMLVEATLDFPEEDIDFLEAADARGKLDGLRRAVDDVLANAQQGAILREGLN
                     VVLVGAPNVGKSSLLNALAGDEVAIVTDIAGTTRDAVRERILIDGVPVHIVDTAGLRE
                     TDDVVERIGIERSRKAVSEADVALVLVDPREGLNEKTRAILDALPPELKRIEIHSKSD
                     LHAHAAGGFGTGAETVIALSAKTGDGLDALKRTLLREAGWQGESEGLFLARTRHVNAL
                     KAAQEELSLAALCGNHQIELFAEHLRLAQVACGEITGEFTADDLLGVIFSRFCIGK"
BASE COUNT      2307 a    2467 c    2721 g    2579 t
ORIGIN
```

```
  Query Match             100.0%;  Score 4374;  DB 1;  Length 10074;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4374;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7237 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 7178

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7177 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 7118

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7117 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 7058

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7057 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 6998

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6997 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 6938

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6937 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 6878

Qy     361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6877 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 6818

Qy     421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6817 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 6758

Qy     481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6757 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 6698

Qy     541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6697 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 6638

Qy     601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6637 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 6578

Qy     661 gcaagtgcgtattcttggctcgttggtggcaatacctttgcacaaaatggatcaggtggt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6577 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 6518

Qy     721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggtttttaccaaca 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6517 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 6458
```

```
Qy    781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6457 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 6398

Qy    841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6397 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 6338

Qy    901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6337 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 6278

Qy    961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6277 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 6218

Qy   1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6217 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 6158

Qy   1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6157 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 6098

Qy   1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacgaa 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6097 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 6038

Qy   1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6037 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 5978

Qy   1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5977 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 5918

Qy   1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5917 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 5858

Qy   1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5857 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 5798

Qy   1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5797 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 5738

Qy   1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5737 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 5678

Qy   1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5677 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 5618

Qy   1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5617 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 5558

Qy   1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggcaat 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5557 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 5498

Qy   1741 aacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5497 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 5438

Qy   1801 acgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcacc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5437 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 5378

Qy   1861 ctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactg 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5377 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 5318

Qy   1921 tttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5317 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 5258

Qy   1981 aaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacattt 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5257 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 5198

Qy   2041 aaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaa 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5197 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 5138

Qy   2101 gtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5137 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 5078

Qy   2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaa 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5077 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 5018

Qy   2221 accattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5017 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 4958

Qy   2281 gatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaatctt 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4957 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 4898

Qy   2341 agtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaaccttt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4897 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 4838

Qy   2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4837 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 4778

Qy   2461 gcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacg 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4777 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 4718

Qy   2521 ctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctcccta 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4717 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 4658

Qy   2581 gccgataaggcagtattccattttgaaagcagccgctttaccggacaaatcagcggcggc 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4657 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 4598

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaatta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4597 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 4538

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcg 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4537 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 4478

Qy   2761 gcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgc 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4477 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 4418

Qy   2821 cgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4417 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 4358

Qy   2881 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4357 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 4298

Qy   2941 cgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaac 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   4297 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 4238

Qy   3001 aataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaac 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4237 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 4178

Qy   3061 aaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcg 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4177 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 4118

Qy   3121 tggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4117 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 4058

Qy   3181 caagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaagacaac 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4057 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC 3998

Qy   3241 gcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagacagaaagc 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3997 GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3938

Qy   3301 gttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcggaggaa 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3937 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA 3878

Qy   3361 gagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaagcggaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3877 GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3818

Qy   3421 acccggccggctaccaccgccttcccccgcgcccgccgcgcccgccgggatttgccgcaa 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3817 ACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3758

Qy   3481 ctgcaaccccaaccgcagccccaaccgcagcgcgacctgatcagccgttatgccaatagc 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3757 CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGC 3698

Qy   3541 ggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacaggacgaattagac 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3697 GGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGAC 3638

Qy   3601 cgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3637 CGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3578

Qy   3661 cactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggt 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3577 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3518

Qy   3721 atgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaa 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3517 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAA 3458

Qy   3781 aacaccttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcggg 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3457 AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGG 3398

Qy   3841 caatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggc 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3397 CAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGC 3338

Qy   3901 agcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcag 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3337 AGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3278

Qy   3961 gcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgc 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3277 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGC 3218

Qy   4021 tatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggcctt 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3217 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTT 3158
```

```
Qy   4081 gcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacac 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3157 GCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACAC 3098

Qy   4141 atttccatcacgccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccga 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3097 ATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGA 3038

Qy   4201 acacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgcagtgcggaatgg 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3037 ACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 2978

Qy   4261 ggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2977 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGC 2918

Qy   4321 ccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2917 CCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 2864
```

```
RESULT   4
AX044034/c
LOCUS       AX044034               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 113 from Patent WO0066791.
ACCESSION   AX044034
VERSION     AX044034.1  GI:11342918
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 113 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.-seq 1: 1 to
                     349980 349980 bases-seq 108: 300001 to 649980 349980
                     bases-seq 109: 600001 to 949980 349980 bases-seq 110:
                     900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
                     349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
                     113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    86473 a  95646 c  85908 g  81953 t
ORIGIN
```

```
  Query Match            100.0%;  Score 4374;  DB 6;  Length 349980;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4374;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  289191 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 289132

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  289131 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 289072

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  289071 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 289012

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 289011 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 288952

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288951 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 288892

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288891 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 288832

Qy    361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288831 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 288772

Qy    421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288771 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 288712

Qy    481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288711 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 288652

Qy    541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288651 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 288592

Qy    601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288591 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 288532

Qy    661 gcaagtgcgtattcttggctcgttggtggcaataccttgcacaaaatggatcaggtggt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288531 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 288472

Qy    721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggtttttttaccaaca 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288471 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTTACCAACA 288412

Qy    781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288411 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 288352

Qy    841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288351 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 288292

Qy    901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288291 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 288232

Qy    961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288231 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 288172

Qy   1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288171 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 288112

Qy   1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288111 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 288052

Qy   1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacgaa 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288051 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 287992

Qy   1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287991 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 287932

Qy   1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287931 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 287872

Qy   1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287871 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 287812

Qy   1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287811 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 287752

Qy   1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287751 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 287692

Qy   1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287691 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 287632

Qy   1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287631 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 287572

Qy   1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287571 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 287512

Qy   1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggcaat 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287511 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 287452

Qy   1741 aacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287451 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 287392

Qy   1801 acgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcacc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287391 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 287332

Qy   1861 ctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactg 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287331 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 287272

Qy   1921 tttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287271 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 287212

Qy   1981 aaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacattt 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287211 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 287152

Qy   2041 aaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaa 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287151 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 287092

Qy   2101 gtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287091 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 287032

Qy   2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaa 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287031 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 286972

Qy   2221 accattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286971 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 286912

Qy   2281 gatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaatctt 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286911 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 286852

Qy   2341 agtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaacctt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286851 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 286792
```

```
Qy   2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286791 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 286732

Qy   2461 gcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacg 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286731 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 286672

Qy   2521 ctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctcccta 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286671 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 286612

Qy   2581 gccgataaggcagtattccatttttgaaagcagccgctttaccggacaaatcagcggcggc 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286611 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 286552

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaatta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286551 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 286492

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcg 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286491 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 286432

Qy   2761 gcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgc 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286431 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 286372

Qy   2821 cgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286371 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 286312

Qy   2881 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286311 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 286252

Qy   2941 cgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaac 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286251 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 286192

Qy   3001 aataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaac 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286191 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 286132

Qy   3061 aaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcg 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286131 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 286072

Qy   3121 tggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286071 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 286012

Qy   3181 caagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaagacaac 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286011 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC 285952

Qy   3241 gcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagacagaaagc 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285951 GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 285892

Qy   3301 gttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcggaggaa 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285891 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA 285832

Qy   3361 gagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaagcggaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285831 GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 285772

Qy   3421 acccggccggctaccaccgccttccccccgcgcccgccgcgcccgccgggatttgccgcaa 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285771 ACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 285712

Qy   3481 ctgcaaccccaaccgcagccccaaccgcagcgcgacctgatcagccgttatgccaatagc 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285711 CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGC 285652

Qy   3541 ggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacaggacgaattagac 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285651 GGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGAC 285592

Qy   3601 cgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285591 CGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 285532

Qy   3661 cactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggt 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285531 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 285472

Qy   3721 atgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaa 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285471 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAA 285412

Qy   3781 aacaccttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcggg 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285411 AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGG 285352

Qy   3841 caatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggc 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285351 CAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGC 285292

Qy   3901 agcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcag 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285291 AGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 285232

Qy   3961 gcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgc 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285231 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGC 285172

Qy   4021 tatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggcctt 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285171 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTT 285112

Qy   4081 gcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacac 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285111 GCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACAC 285052

Qy   4141 atttccatcacgccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccga 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 285051 ATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGA 284992

Qy   4201 acacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgcagtgcggaatgg 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 284991 ACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 284932

Qy   4261 ggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 284931 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGC 284872

Qy   4321 ccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 284871 CCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 284818
```

```
RESULT   5
NME242535
LOCUS       NME242535               4374 bp    DNA     linear   BCT 16-JAN-2002
DEFINITION  Neisseria meningitidis serogroup B app gene, strain B:15:P1.16.
ACCESSION   AJ242535
VERSION     AJ242535.2  GI:6977940
KEYWORDS    APP gene; app protein.
SOURCE      Neisseria meningitidis serogroup B.
  ORGANISM  Neisseria meningitidis serogroup B
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
```

```
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (sites)
  AUTHORS   Ala Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1999) Ala Aldeen D.A., Microbiology, University
            of Nottingham, Division of Microbiology, SCLS, University Hospital,
            Nottingham, NG7 2UH, UNITED KINGDOM
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 4374)
  AUTHORS   Ala Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) Ala Aldeen D.A., Microbiology, University
            of Nottingham, Division of Microbiology, SCLS, University Hospital,
            Nottingham, NG7 2UH, UNITED KINGDOM
COMMENT     On Feb 14, 2000 this sequence version replaced gi:4938471.
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="Neisseria meningitidis serogroup B"
                     /strain="B:15:P1.16"
                     /isolate="SD"
                     /db_xref="taxon:491"
     gene            1. .4374
                     /gene="app"
     CDS             1. .4374
                     /gene="app"
                     /codon_start=1
                     /transl_table=11
                     /product="App protein"
                     /protein_id="CAB43832.2"
                     /db_xref="GI:6977941"
                     /db_xref="SPTREMBL:Q9X7H1"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDAHSVFYEPRQNGKYSFND
                     DNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINQGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGSLSANGDTRYTVSHYATQNG
                     NLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHNAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFR
                     FMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFT
                     LQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALI
                     AAGRDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATT
                     AFPRARRARRDLPQLQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFA
                     EDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENT
                     FDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVLHYGIQ
                     ARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA
                     QHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAA
                     AAKGPQLEAQHSAGIKLGYRW"
BASE COUNT      1265 a   1163 c   1066 g    880 t
ORIGIN
```

```
  Query Match             99.8%;  Score 4366;  DB 1;  Length 4374;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 4369;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy     421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 480

Qy     481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 540

Qy     541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600

Qy     601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660

Qy     661 gcaagtgcgtattcttggctcgttggtggcaataccttgcacaaaatggatcaggtggt 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy     721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggtttttaccaaca 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840

Qy     841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

Qy     901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATGCCCATTCAGTA 960

Qy     961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 1020

Qy    1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080

Qy    1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140

Qy    1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttccttttattgacgaa 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200

Qy   1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 1260

Qy   1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320

Qy   1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380

Qy   1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440

Qy   1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500

Qy   1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560

Qy   1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1620

Qy   1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680

Qy   1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggcaat 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 1740

Qy   1741 aacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800

Qy   1801 acgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcacc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860

Qy   1861 ctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactg 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920

Qy   1921 tttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 1980

Qy   1981 aaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacattt 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 2040

Qy   2041 aaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaa 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 2100

Qy   2101 gtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 2160

Qy   2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaa 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 2220

Qy   2221 accattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 2280

Qy   2281 gatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaatctt 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAGTCTT 2340

Qy   2341 agtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaacctt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACTACGCCACCCAAAACGGCAACCTT 2400

Qy   2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460

Qy   2461 gcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacg 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520

Qy   2521 ctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctcccta 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580

Qy   2581 gccgataaggcagtattccattttgaaagcagccgctttaccggacaaatcagcggcggc 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGC 2640

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaatta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCATCAGGCACGGAATTA 2700

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcg 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACAATGCG 2760

Qy   2761 gcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgc 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGC 2820

Qy   2821 cgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 2880

Qy   2881 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2940

Qy   2941 cgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaac 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000

Qy   3001 aataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaac 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 3060

Qy   3061 aaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcg 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 3120

Qy   3121 tggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3180

Qy   3181 caagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaagacaac 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC 3240

Qy   3241 gcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagacagaaagc 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3300

Qy   3301 gttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcggaggaa 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA 3360
```

```
Qy    3361 gagaaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaagcggaa 3420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3361 GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3420

Qy    3421 acccggccggctaccaccgccttccccccgcgcccgccgcgcccgccgggatttgccgcaa 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3421 ACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3480

Qy    3481 ctgcaaccccaaccgcagccccaaccgcagcgcgacctgatcagccgttatgccaatagc 3540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3481 CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGC 3540

Qy    3541 ggtttgagtgaatttccgccacgctcaacagcgttttcgccgtacaggacgaattagac 3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3541 GGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGAC 3600

Qy    3601 cgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaa 3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3601 CGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3660

Qy    3661 cactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggt 3720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3661 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3720

Qy    3721 atgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaa 3780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3721 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAA 3780

Qy    3781 aacaccttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcggg 3840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3781 AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGG 3840

Qy    3841 caatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggc 3900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3841 CAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGC 3900

Qy    3901 agcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcag 3960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3901 AGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3960

Qy    3961 gcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgc 4020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3961 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGC 4020

Qy    4021 tatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggcctt 4080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4021 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTT 4080

Qy    4081 gcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacac 4140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4081 GCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACAC 4140

Qy    4141 atttccatcacgccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccga 4200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4141 ATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGA 4200

Qy    4201 acacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgcagtgcggaatgg 4260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4201 ACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 4260

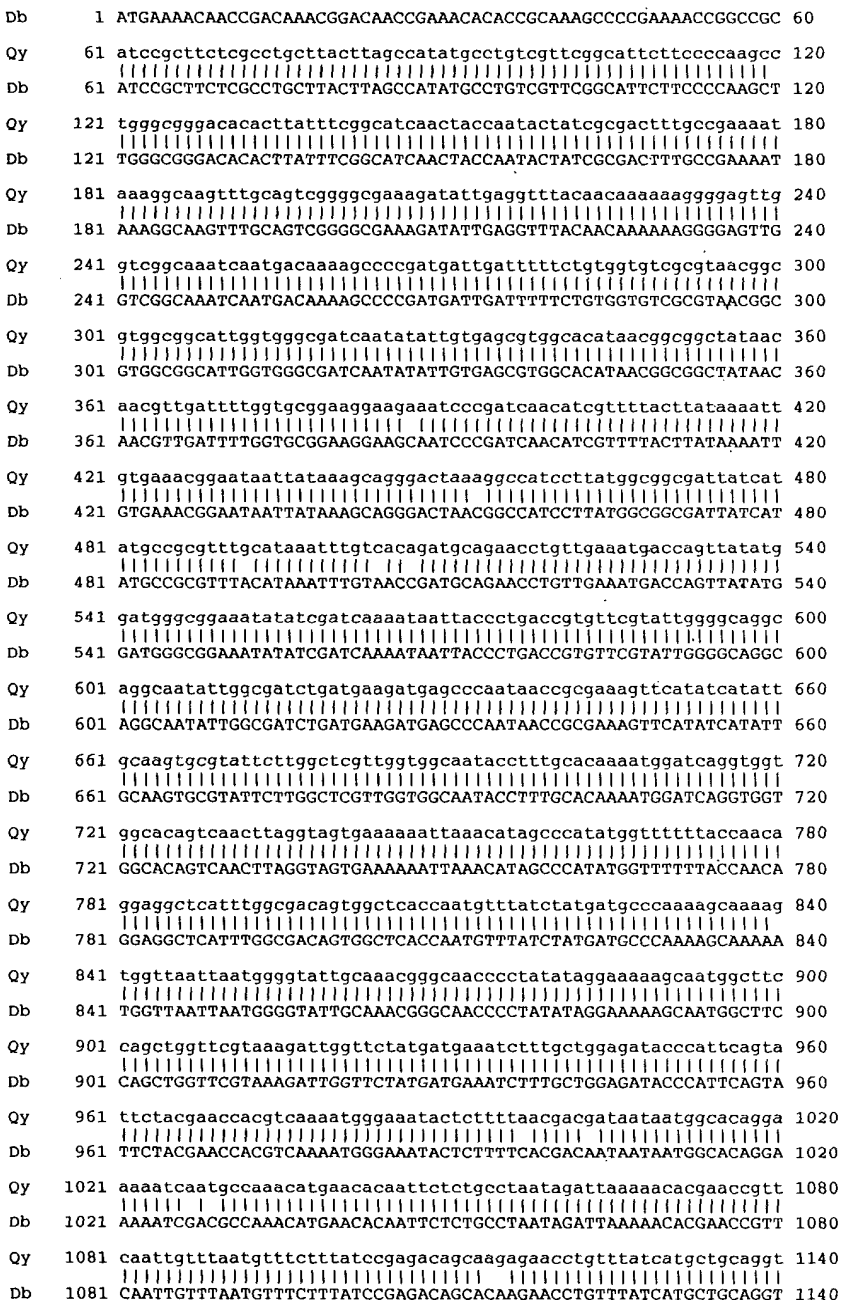
Qy    4261 ggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggc 4320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4261 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGC 4320

Qy    4321 ccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4321 CCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4374


RESULT   6
NME296277
LOCUS       NME296277              4386 bp    DNA     linear   BCT 14-JAN-2002
DEFINITION  Neisseria meningitidis app gene for adhesion penetration protein,
            strain Z4181.
ACCESSION   AJ296277
VERSION     AJ296277.1  GI:11071864
KEYWORDS    adhesion penetration protein; APP gene.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (bases 1 to 4386)
  AUTHORS   Ala-Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2000) Ala-Aldeen D.A., Microbiology, University
            of Nottingham, Molecular Bacteriology and Immunology Group,
            University Hospital, Nottingham NG7 2UH, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .4386
                     /organism="Neisseria meningitidis"
                     /strain="Z4181"
                     /db_xref="taxon:487"
                     /country="Mali"
                     /note="serogroup C"
     gene            1. .4365
                     /gene="app"
     CDS             1. .4365
                     /gene="app"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion penetration protein"
                     /protein_id="CAC14671.1"
                     /db_xref="GI:11071865"
                     /db_xref="SPTREMBL:Q9F3Z4"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFTYKIVKRNNYKAGTNGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVFYEPRQNGKYSFHD
                     NNNGTGKIDAKHEHNSLPNRLKTRTVQLFNVSLSETAQEPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINRGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGKVILDQQADENNKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTNTDISGNVSLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNG
                     DLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHSESSRFTGQISGSKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHDAAGAQTGSATDAPRRRSRRSLLSVTPPASVESRFNTLTVNGKLNGQGTFRFMS
                     ELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFTLQN
                     EHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALIAAG
                     RDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATTAFP
                     RARRARRDLPQLQPQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEGR
                     RNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDD
                     GIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGSKIRRRVLHYGIQARY
                     RAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
                     SITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAK
                     GPQLEAQHSAGIKLGYRW"
BASE COUNT     1270 a   1167 c   1066 g    883 t
ORIGIN


  Query Match             97.5%;  Score 4266;  DB 1;  Length 4386;
  Best Local Similarity   98.7%;  Pred. No. 0;
  Matches 4315;  Conservative    0;  Mismatches   50;  Indels    9;  Gaps    1;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```


Db 1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60
Qy 61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
Db 61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120
Qy 121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
Db 121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180
Qy 181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
Db 181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240
Qy 241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
Db 241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300
Qy 301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
Db 301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360
Qy 361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420
Db 361 AACGTTGATTTTGGTGCGGAAGGAAGCAATCCCGATCAACATCGTTTTACTTATAAAATT 420
Qy 421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480
Db 421 GTGAAACGGAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCAT 480
Qy 481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
Db 481 ATGCCGCGTTTACATAAATTTGTAACCGATGCAGAACCTGTTGAAATGACCAGTTATATG 540
Qy 541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
Db 541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600
Qy 601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
Db 601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660
Qy 661 gcaagtgcgtattcttggctcgttggtggcaatacctttgcacaaaatggatcaggtggt 720
Db 661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720
Qy 721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggtttttaccaaca 780
Db 721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780
Qy 781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
Db 781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAA 840
Qy 841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
Db 841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900
Qy 901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
Db 901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960
Qy 961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
Db 961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTCACGACAATAATAATGGCACAGGA 1020
Qy 1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
Db 1021 AAAATCGACGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080
Qy 1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db 1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCACAAGAACCTGTTTATCATGCTGCAGGT 1140
Qy 1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacgaa 1200
Db 1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200
Qy 1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
Db 1201 GGAAAAGGCGAATTGATACTTACCAGTAACATCAACCGAGGTGCTGGAGGATTATATTTC 1260
Qy 1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
Db 1261 CAAGGAGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
Qy 1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
Db 1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTCAATGGCGTGGCAAACGACCGCCTGTCC 1380
Qy 1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
Db 1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440
Qy 1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
Db 1441 GTGGGCGACGGTAAAGTTATTTTAGATCAACAAGCAGATGAAAATAATAAAAAACAAGCC 1500
Qy 1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
Db 1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560
Qy 1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
Db 1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
Qy 1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
Db 1621 TCACTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
Qy 1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggcaat 1740
Db 1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT 1740
Qy 1741 aacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
Db 1741 AACAACAGCTTGGATAGCAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
Qy 1801 acgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcacc 1860
Db 1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCTGCAGAAGACCGCACC 1860
Qy 1861 ctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactg 1920
Db 1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
Qy 1921 tttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaa 1980
Db 1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAA 1980
Qy 1981 aaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacattt 2040
Db 1981 AAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTT 2040
Qy 2041 aaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaa 2100
Db 2041 AAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAA 2100
Qy 2101 gtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcat 2160
Db 2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 2160
Qy 2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaa 2220
Db 2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAA 2220

```
Qy   2221 accattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtc 2280
          ||||||||||||||||||||||||||||||||||| || ||||||||||||||||||||||
Db   2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGACCAATACCGACATCAGCGGCAATGTC 2280

Qy   2281 gatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaatcctt 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AGCCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTT 2340

Qy   2341 agtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaacctt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 AGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCGACCTT 2400

Qy   2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460

Qy   2461 gcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacg 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520

Qy   2521 ctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctccccta 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580

Qy   2581 gccgataaggcagtattccattttgaaagcagccgctttaccggacaaatcagcggcggc 2640
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||| ||
Db   2581 GCCGATAAGGCAGTATTCCATTCTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCAGC 2640

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaatta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcg 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCG 2760

Qy   2761 gcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgc 2820
          |||||||||||||||||||||||||||||||||||||||||         ||||||||||||
Db   2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGC---------CGCCGTTCGCGC 2811

Qy   2821 cgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacg 2880
          |||||||||||||||||||| ||  ||||||||||||||||||||||||||||||||||||
Db   2812 CGTTCCCTATTATCCGTTACACCTCCGGCTTCGGTAGAATCCCGTTTCAACACGCTGACG 2871

Qy   2881 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2872 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2931

Qy   2941 cgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaac 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2932 CGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 2991

Qy   3001 aataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaac 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2992 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 3051

Qy   3061 aaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcg 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3052 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCG 3111

Qy   3121 tggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3112 TGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3171

Qy   3181 caagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaagacaac 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3172 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC 3231

Qy   3241 gcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagacagaaagc 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3232 GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3291

Qy   3301 gttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcggaggaa 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3292 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA 3351

Qy   3361 gagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaagcggaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3352 GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3411

Qy   3421 acccggccggctaccaccgccttcccccgcgcccgccgcgcccgccgggatttgccgcaa 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3412 ACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3471

Qy   3481 ctgcaaccccaaccgcagccccaaccgcagcgcgacctgatcagccgttatgccaatagc 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3472 CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGC 3531

Qy   3541 ggtttgagtgaatttttccgccacgctcaacagcgttttcgccgtacaggacgaattagac 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3532 GGTTTGAGTGAATTTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGAC 3591

Qy   3601 cgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaa 3660
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   3592 CGCGTATTTGCCGAAGGCCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3651

Qy   3661 cactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggt 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3652 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3711

Qy   3721 atgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaa 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3712 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAA 3771

Qy   3781 aacaccttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcggg 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3772 AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGG 3831

Qy   3841 caatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggc 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3832 CAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGC 3891

Qy   3901 agcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcag 3960
          || |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3892 AGTCTTTCAGACGGCATCGGAAGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3951

Qy   3961 gcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgc 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3952 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGC 4011

Qy   4021 tatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggcctt 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4012 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTT 4071

Qy   4081 gcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacac 4140
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4072 GCGTTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACAC 4131

Qy   4141 atttccatcacgccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccga 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4132 ATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGA 4191

Qy   4201 acacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgcagtgcggaatgg 4260
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4192 ACGCGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 4251

Qy   4261 ggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4252 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGC 4311

Qy   4321 ccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4312 CCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4365

RESULT   7
```

```
A96620
LOCUS       A96620                  4407 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 653 from Patent WO9924578.
ACCESSION   A96620
VERSION     A96620.1  GI:6780202
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4407)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 653 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4407
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     1273 a   1169 c   1094 g    871 t
ORIGIN
```

Query Match 90.4%; Score 3954.8; DB 6; Length 4407;
Best Local Similarity 94.0%; Pred. No. 0;
Matches 4153; Conservative 0; Mismatches 212; Indels 51; Gaps 2;

```
Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAAAACCGGCCGC 60

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
Db      61 ATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCCCCAAGCC 120

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
Db     121 CGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcgggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
Db     241 GTCGGCAAATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
Db     301 GTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420
Db     361 AATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATT 420

Qy     421 gtgaaacggaataattattaaagcagggactaaaggccatccttatggcggcgattatcat 480
Db     421 GTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCAT 480

Qy     481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
Db     481 ATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATG 540

Qy     541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
Db     541 GATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGC 600

Qy     601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
Db     601 AGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATT 660

Qy     661 gcaagtgcgtattcttggctcgttggtggcaatacctttgcacaaaatggatcaggtggt 720
Db     661 GCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy     721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggtttttaccaaca 780
Db     721 GGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780

Qy     781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
Db     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840

Qy     841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
Db     841 TGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

Qy     901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
Db     901 CAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960

Qy     961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
Db     961 TTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGA 1020

Qy    1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
Db    1021 AAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTT 1080

Qy    1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db    1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140

Qy    1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacgaa 1200
Db    1141 GGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAA 1200

Qy    1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
Db    1201 GGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 1260

Qy    1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
Db    1261 GAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320

Qy    1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
Db    1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380

Qy    1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
Db    1381 AAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGC 1440

Qy    1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
Db    1441 GTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCC 1500

Qy    1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
Db    1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAG 1560

Qy    1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
Db    1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620

Qy    1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
Db    1621 TCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680

Qy    1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggcaat 1740
Db    1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAAT 1740

Qy    1741 aacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagat 1800
Db    1741 AACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
```

```
Qy   1801 acgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcacc 1860
Db   1801 GCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACT 1860

Qy   1861 ctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactg 1920
Db   1861 TTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTG 1920

Qy   1921 tttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaa 1980
Db   1921 TTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAA 1980

Qy   1981 aaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacattt 2040
Db   1981 ATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTT 2040

Qy   2041 aaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaa 2100
Db   2041 AAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAA 2100

Qy   2101 gtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcat 2160
Db   2101 GTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCAT 2160

Qy   2161 caaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaa 2220
Db   2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAA 2220

Qy   2221 accattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtc 2280
Db   2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTC 2280

Qy   2281 gatcttgccgatcacgctcatttaaatctcacaggcttgccacactcaacggcaatctt 2340
Db   2281 AGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTT 2340

Qy   2341 agtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaacctt 2400
Db   2341 AGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTC 2400

Qy   2401 agcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcg 2460
Db   2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460

Qy   2461 gcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacg 2520
Db   2461 GCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACG 2520

Qy   2521 ctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctcccta 2580
Db   2521 CTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTA 2580

Qy   2581 gccgataaggcagtattccatttgaaagcagccgctttaccggacaaatcagcggcggc 2640
Db   2581 GCCGATAAGGCAGTATTCCATTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGC 2640

Qy   2641 aaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaatta 2700
Db   2641 AAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTA 2700

Qy   2701 ggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcg 2760
Db   2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCG 2760

Qy   2761 gcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgc 2820
Db   2761 GCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGC---------CGCCGTTCGCGC 2811

Qy   2821 cgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacg 2880
Db   2812 CGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACG 2871

Qy   2881 gtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctac 2940
Db   2872 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTAC 2931

Qy   2941 cgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaac 3000
Db   2932 CGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAAC 2991

Qy   3001 aataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaac 3060
Db   2992 AATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAAC 3051

Qy   3061 aaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcg 3120
Db   3052 ACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCA 3111

Qy   3121 tggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaa 3180
Db   3112 TGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3171

Qy   3181 caagagctttccgacaaactcggcaaggcagaa---------------------------- 3213
Db   3172 CAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAA 3231

Qy   3214 ---------------gccaaaaaacaggcggaaaaagacaacgcgcaaagccttgacgcg 3258
Db   3232 CAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCG 3291

Qy   3259 ctgattgcggccgggcgcgatgccgtcgaaaagacagaaagcgttgccgaaccggcccgg 3318
Db   3292 CTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGG 3351

Qy   3319 caggcaggcggggaaaatgtcggcattatgcaggcggaggaagagaaaaaacgggtgcag 3378
Db   3352 CAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3411

Qy   3379 gcggataaagacaccgccttggcgaaacagcgcgaagcggaaacccggccggctaccacc 3438
Db   3412 GCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACC 3471

Qy   3439 gccttcccccgcgcccgccgcgcccgccgggatttgccgcaactgcaaccccaaccgcag 3498
Db   3472 GCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAA 3531

Qy   3499 ccccaaccgcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaattttcc 3558
Db   3532 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCC 3591

Qy   3559 gccacgctcaacagcgttttcgccgtacaggacgaattagaccgcgtatttgccgaagac 3618
Db   3592 GCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGAC 3651

Qy   3619 cgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcgcaagat 3678
Db   3652 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGAT 3711

Qy   3679 ttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggc 3738
Db   3712 TTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGC 3771

Qy   3739 agcgggcgcgtcggcatcctgttttcgcacaaccggaccgaaaacaccttcgacgacggc 3798
Db   3772 AGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGC 3831

Qy   3799 atcggcaactcggcacggcttgcccacggcgccgttttcgggcaatacggcatcgacagg 3858
Db   3832 ATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGG 3891

Qy   3859 ttctacatcggcatcagcgcgggcgcgggttttagcagcggcagcctttcagacggcatc 3918
Db   3892 TTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATC 3951

Qy   3919 ggaggcaaaatccgccgccgcgtgctgcattacggcattcaggcacgataccgcgccggt 3978
```

```
Db    3952 AGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGT 4011

Qy    3979 ttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaaaaagcg 4038

Db    4012 TTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCG 4071

Qy    4039 gattaccgctacgaaaacgtcaatatcgccacccccggccttgcattcaaccgctaccgc 4098

Db    4072 GATTACCGATACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGC 4131

Qy    4099 gcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacgccttat 4158

Db    4132 GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTAT 4191

Qy    4159 ttgagcctgtcctataccgatgccgcttcgggcaaagtccgaacacgcgtcaataccgcc 4218

Db    4192 TTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCC 4251

Qy    4219 gtattggctcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatc 4278

Db    4252 GTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATC 4311

Qy    4279 aaaggtttcacgctgtccctccacgctgccgccgccaaaggcccgcaactggaagcgcaa 4338

Db    4312 AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAG 4371

Qy    4339 cacagcgcgggcatcaaattaggctaccgctggtaa 4374

Db    4372 CACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4407
```

```
RESULT   8
NME296276
LOCUS       NME296276              4350 bp    DNA     linear   BCT 14-JAN-2002
DEFINITION  Neisseria meningitidis app gene for adhesion penetration protein,
            strain Z3515.
ACCESSION   AJ296276
VERSION     AJ296276.1  GI:11071862
KEYWORDS    adhesion penetration protein; APP gene.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (bases 1 to 4350)
  AUTHORS   Ala-Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2000) Ala-Aldeen D.A., Microbiology, University
            of Nottingham, Molecular Bacteriology and Immunology Group,
            University Hospital, Nottingham NG7 2UH, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .4350
                     /organism="Neisseria meningitidis"
                     /strain="Z3515"
                     /db_xref="taxon:487"
                     /country="Saudi Arabia"
                     /note="serogroup A"
     gene            1. .4350
                     /gene="app"
     CDS             1. .4350
                     /gene="app"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion penetration protein"
                     /protein_id="CAC14670.1"
                     /db_xref="GI:11071863"
                     /db_xref="SPTREMBL:Q9F3Z5"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHM
                     PRLHKFVTDAEPVEMTSDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGA
                     WLIGGNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLL
                     NGVLQTGYPYSGRENGFQLIRKDWFYDDIYRGDTHTVFFEPRSNGHFSFTSNNNGTGT
                     VTETNEKVSNPKLEVQTVRLFDESLNETDKEPVYAAGGVNQYRPRLNNGENLSFIDYG
                     NGKLILSNNINQGAGGLYFEGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRL
                     SKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVSGRGTVQLNA
                     DNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNATTTSTVTITGNESI
                     TQPSGKNINRLNYSKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLSGGTNLNG
                     NITQTNGKLFFSGRPTPHAYNHLGSGWSKMEGIPQGEIVWDNDWINRTFKAENFHIQG
                     GQAVISRNVAKVEGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEKTITDDKV
                     IASLTKTDISGNVSLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNGNLSLVGN
                     AQATFNQATLNGNTSASGNASFNLSNNAAQNGSLTLSDNAKANVSHSALNGNVSLADK
                     AVFHFENSRFTGQLSGSKDTALHLKDSEWTLPSGTELGNLNLDNATITLNSAYRHDAA
                     GAQTGSVSDTPRRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFRFMSELFGYRS
                     DKLKLAESSEGTYTLAVNNTGNEPVSLDQLTVVEGKDNKPLSENLNFTLQNEHVDAGA
                     WRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALIAAGRDAAEKT
                     ESVAEPARQAGGENVGIMQAEEEKKRVQADKDSALAKQREAETRPATTAFPRARRARR
                     DLPQPQPQPQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVW
                     TSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNS
                     ARLAHGAVFGQYGIGRFDIGISTGAGFSSGSLSDGIGGKIRRRVLHYGIQARYRAGFG
                     GFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPY
                     LSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGPQLE
                     AQHSAGIKLGYRW"
BASE COUNT     1236 a   1185 c   1073 g    856 t
ORIGIN
```

```
  Query Match             81.5%;  Score 3566.8;  DB 1;  Length 4350;
  Best Local Similarity   89.7%;  Pred. No. 0;
  Matches 3936;  Conservative    0;  Mismatches  402;  Indels   48;  Gaps    8;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60

Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120

Db     61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180

Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240

Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300

Db    241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360

Db    301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy    361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420

Db    361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAGCACCGTTTTTCTTACCAAATT 420

Qy    421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480

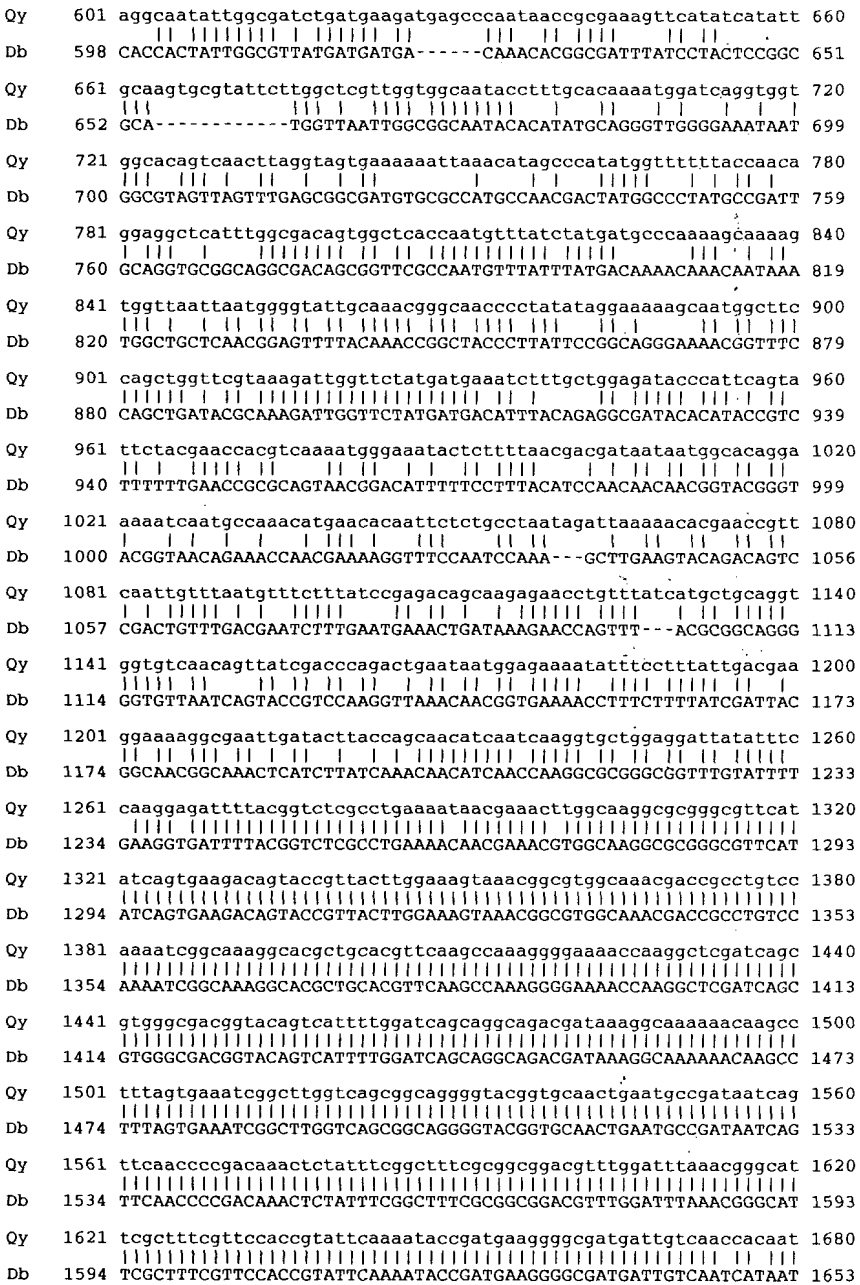
Db    421 GTGAAAAGAAATAATTATAAGCCTGACAATTCA---CACCCTTACAACGGCGATTACCAT 477

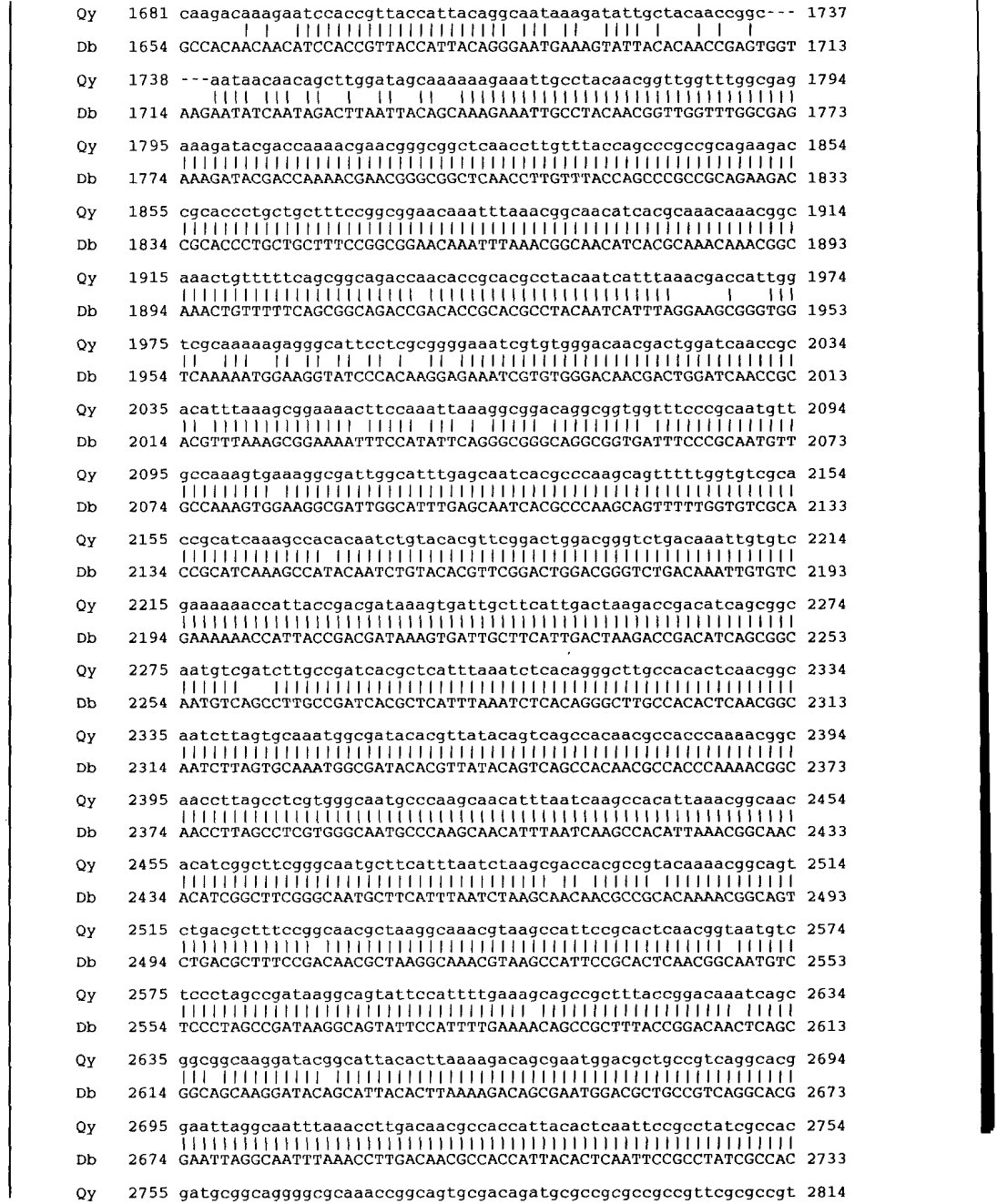
Qy    481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540

Db    478 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATG 537

Qy    541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600

Db    538 AGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGA 597
```





```
Db   2734 GATGCTGCAGGCGCGCAAACCGGCAGTGTGTCAGACACGCCGCCGCCGCCGTTCCGCGCCG- 2792

Qy   2815 tcgcgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacg 2874
Db   2793 --------TTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 2844

Qy   2875 ctgacggtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttc 2934
Db   2845 CTGACGGTAAACGGCAAATTGAACGGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTC 2904

Qy   2935 ggctaccgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcg 2994
Db   2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCG 2964

Qy   2995 gtcaacaataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaa 3054
Db   2965 GTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAA 3024

Qy   3055 gacaacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgcc 3114
Db   3025 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCC 3084

Qy   3115 ggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtc 3174
Db   3085 GGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC 3144

Qy   3175 aaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaa 3234
Db   3145 AAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 3204

Qy   3235 gacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagaca 3294
Db   3205 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACA 3264

Qy   3295 gaaagcgttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcg 3354
Db   3265 GAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3324

Qy   3355 gaggaagagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaa 3414
Db   3325 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACAGCGCCTTGGCGAAACAGCGCGAA 3384

Qy   3415 gcggaaacccggccggctaccaccgccttcccccgcgcccgccgcgcccgccgggatttg 3474
Db   3385 GCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCCCCGGGATTTG 3444

Qy   3475 ccgcaactgcaaccccaaccg------cagccccaaccgcagcgcgacctgatcagccgt 3528
Db   3445 CCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATCAGCCGT 3504

Qy   3529 tatgccaatagcggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacag 3588
Db   3505 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAG 3564

Qy   3589 gacgaattagaccgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatc 3648
Db   3565 GACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATC 3624

Qy   3649 cgggacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctg 3708
Db   3625 CGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 3684

Qy   3709 cgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcac 3768
Db   3685 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCAC 3744

Qy   3769 aaccggaccgaaaacaccttcgacgacggcatcggcaactcggcacggcttgcccacggc 3828
Db   3745 AACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3804

Qy   3829 gccgttttcgggcaatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggt 3888
Db   3805 GCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGT 3864

Qy   3889 tttagcagcggcagcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcat 3948
Db   3865 TTTAGCAGCGGCAGTCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCAT 3924

Qy   3949 tacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatc 4008
Db   3925 TACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATC 3984

Qy   4009 ggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgcc 4068
Db   3985 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCC 4044

Qy   4069 accccggccttgcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaa 4128
Db   4045 ACCCCCGGTCTTGCGTTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAA 4104

Qy   4129 ccggcgcaacacatttccatcacgccttatttgagcctgtcctataccgatgccgcttcg 4188
Db   4105 CCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCG 4164

Qy   4189 ggcaaagtccgaacacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgc 4248
Db   4165 GGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGC 4224

Qy   4249 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgcc 4308
Db   4225 AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCC 4284

Qy   4309 gccgccaaaggcccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4368
Db   4285 GCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGC 4344

Qy   4369 tggtaa 4374
Db   4345 TGGTAA 4350
```

```
RESULT   9
NMA2Z2491
LOCUS       NMA2Z2491              349061 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 2/7.
ACCESSION   AL162753 AL157959
VERSION     AL162753.2  GI:7379120
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349061)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 349061)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
```

```
source          1. .349061
                /organism="Neisseria meningitidis Z2491"
                /strain="Z2491"
                /db_xref="taxon:122587"
                /note="serogroup: A"
repeat_unit     complement(24. .206)
                /note="ATR repeat; hmmfs hit to HMM ATR (1 - 183), score:
                310.39"
                /label=ATR
RBS             209. .212
gene            220. .681
                /gene="NMA0368"
CDS             220. .681
                /gene="NMA0368"
                /note="NMA0368, probable integral membrane protein, len:
                153 aa; contains four probable transmembrane domains"
                /codon_start=1
                /transl_table=11
                /product="putative integral membrane protein"
                /protein_id="CAB83669.1"
                /db_xref="GI:7379121"
                /db_xref="SPTREMBL:Q9JWH7"
                /translation="MQEQNRKSSFPIVMLLVSVALWIASLSNVAFYLGNHGSMEGLTV
                LILGSIFASLDIRYCAVYANYVWLAAIVLLALRKKVVPVHAAFWGLALVAFSVKAVYV
                DEAGNTSDIVRYGAGFYLWYAAFAVASIGTFAGKNKERKAASAADGTKNDV"
gene            671. .1492
                /gene="hemK"
CDS             671. .1492
                /gene="hemK"
                /note="NMA0369, hemK, HemK protein, len: 273 aa; similar
                to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
                scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
                PS00092 N-6 Adenine-specific DNA methylases signature"
                /codon_start=1
                /transl_table=11
                /product="HemK protein"
                /protein_id="CAB83670.1"
                /db_xref="GI:7379122"
                /db_xref="SPTREMBL:Q9JWH6"
                /translation="MTFDKWLGLSKLPKNEARMLLQYVSEYTRVQLLTRGGEEMPDEV
                RQRADRLAQRRLNGEPVAYILGAREFYGRRFTVNPSVLIPRPETEHLVEAVLARLPEN
                GRVWDLGTGSGAVAVTVALERPDAFVRASDISPPALETARKNAADLGARVEFAYGSWF
                DTDMPSEGKWDIIVSNPPYIENGDKHLSQGDLRFEPQIALTDFSDGLSCIRTLAQGAP
                DRLAEGGFLLLEHGFDQGAAVRGVLAENGFSGVETLPDLAGLDRVTLGKYMKHLK"
misc_feature    831. .840
                /gene="hemK"
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    1162. .1171
                /gene="hemK"
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    1187. .1207
                /gene="hemK"
                /note="PS00092 N-6 Adenine-specific DNA methylases
                signature"
misc_feature    complement(1279. .1288)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
RBS             1565. .1568
gene            1578. .2966
                /gene="NMA0370"
CDS             1578. .2966
                /gene="NMA0370"
                /note="NMA0370, probable integral membrane protein, len:
                462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
                P44640 hypothetical protein HI0325 (450 aa), fasta scores;
                E(): 0, 49.8% identity in 464 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="putative integral membrane protein"
                /protein_id="CAB83671.1"
                /db_xref="GI:7379123"
                /db_xref="SPTREMBL:Q9JWH5"
                /translation="MNAVVVAVIVMLVLSLSRVHVVLSLTIGAFVGGAVAGMPLQNIA
                DAAGQVSQAGIIPVFNKGLEGGAKIALSYAMLGAFAMAITHSGLPQQLAGAVVRKLNR
                GGMPDSVRSGEGVVKWLLLSIILVMGMMSQNIIPIHIAFIPMIVPPLLLVFNRLKIDR
                RLIACVITFGLVTTYMFLPYGFGAIFLNEILLGNIHSAAPQLDVKNINVMAAMAIPAL
                GMLAGLLLAFVHYRKPRLYQSNNADTAGNADAANRPQPSAYRSLAAAVAIAVCFAIQL
                MYEDSLVLGAMLGFAVFMMLGVINRDKANDVFGEGIKMMAMVGFIMIAAQGFAAVMNA
                TGHIQPLVESSMAIFGNSKGMAALAMLVVGLLVTMGIGSSFSTLPIIAAIYVPLCVGL
                GFSPLATVAIVGTAGALGDAGSPASDSTLGPTMGLNADGQHDHIRDSVIPTFIHYNIP
                LLIAGWIAAMVL"
gene            2967. .3191
                /gene="slyX"
CDS             2967. .3191
                /gene="slyX"
                /note="NMA0371, slyX, SLYX protein homolog, len: 74 aa;
                simlar to SLYX_HAEIN P44759 SLYX protein homolog (73 aa),
                fasta scores; E(): 0.0017, 33.8% identity in 74 aa
                overlap, and SLYX_ECOLI P30857 SLYX protein (72 aa), fasta
                scores; E(): 0.24, 32.4% identity in 68 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="SLYX protein homolog"
                /protein_id="CAB83672.1"
                /db_xref="GI:7379124"
                /db_xref="SPTREMBL:Q9JWH4"
                /translation="MDAVQEFEHRITELEIQSALQEDVIAGLNAMVAELRQTLDLQQA
                QLRLLYQKMQDRNPDAQEPYSLRDEIPPHY"
gene            complement(3271. .3423)
                /gene="NMA0372"
CDS             complement(3271. .3423)
                /gene="NMA0372"
                /note="NMA0372, unknown, questionable CDS, len: 50 aa"
                /codon_start=1
                /transl_table=11
                /product="very hypothetical protein NMA0372"
                /protein_id="CAB83673.1"
                /db_xref="GI:7379125"
                /db_xref="SPTREMBL:Q9JWH3"
                /translation="MFGRLSPLGSDGFFLAVYEYPVGFICLQGKTAYFRSGGKRFHRI
                KGILSD"
misc_feature    3562. .3571
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(3586. .4356)
                /gene="thiF"
CDS             complement(3586. .4356)
                /gene="thiF"
                /note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
                similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
                fasta scores; E(): 0, 43.1% identity in 246 aa overlap,
                and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
                protein. (249 aa), fasta scores; E(): 0, 43.9% identity in
                244 aa overlap (note that N.m. does not have orthologs of
                any other molybdopterin biosynthesis proteins). Contains
                Pfam match to entry PF00899 ThiF_family, ThiF family"
                /codon_start=1
                /transl_table=11
                /product="ThiF protein"
                /protein_id="CAB83674.1"
                /db_xref="GI:7379126"
                /db_xref="SPTREMBL:Q9JWH2"
                /translation="MTTTEHDNDDAFLLRYSRHILLDEIGIEGQQKLSAAHILVVGCG
                GLGAAALPYLAASGIGTLTIADSDTVELHNLQRQVAFDEGDVGKLKTEALADRLRHIN
                HTVDVRTINEKLDGCRLTGLVQAADIVLDCCDNYATRQAVNRACVQAKTPLVSGAAVR
                FEGQLAVYRPDLPDSPCYACLFDGGSASDGICSLFGVFSPLVGIIGSTQAAEALKILL
                DAGEPSHGRLAVYRALEGGWQYFDLPRNPECPVCGAER"
misc_feature    3787. .3796
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    complement(3847. .4257)
                /gene="thiF"
                /note="Pfam match to entry PF00899 ThiF_family, ThiF
                family, score 186.60, E-value 4.1e-52"
```

```
     misc_feature    complement(4418. .4427)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            4471. .7224
                     /gene="ppc"
     CDS             4471. .7224
                     /gene="ppc"
                     /EC_number="4.1.1.31"
                     /note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
                     917 aa; simlar to many e.g. CAPP_RHOPA O32483
                     phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
                     fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
                     Contains 2x Pfam match to entry PF00311 PEPcase,
                     Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
                     site motif A (P-loop), and PS00393 Phosphoenolpyruvate
                     carboxylase active site 2"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoenolpyruvate carboxylase"
                     /protein_id="CAB83675.1"
                     /db_xref="GI:7379127"
                     /db_xref="SPTREMBL:Q9JWH1"

  Query Match             81.5%;  Score 3566.8;  DB 1;  Length 349061;
  Best Local Similarity   89.7%;  Pred. No. 0;
  Matches 3936;  Conservative    0;  Mismatches  402;  Indels   48;  Gaps    8;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94066 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 94125

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94126 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 94185

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94186 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 94245

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94246 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 94305

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94306 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 94365

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94366 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 94425

Qy    361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaatt 420
          |||||||||||||||||||||||||||||||||||| || |||||| ||||  |||||
Db  94426 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAGCACCGTTTTTCTTACCAAATT 94485

Qy    421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480
          |||||| |||||||||||  | |  | |  |   || |||||  |||||||||| |||
Db  94486 GTGAAAAGAAATAATTATAAGCCTGACAATTCA---CACCCTTACAACGGCGATTACCAT 94542

Qy    481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
          |||||||||||||||||||||||||||||||||||| |||||||| ||| | |||
Db  94543 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATG 94602

Qy    541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
            |||   |  |||  |||| || | || || || || ||||| || || ||  ||||
Db  94603 AGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGA 94662

Qy    601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
            || |||||||| | |||||| ||      |||  || ||||    || ||
Db  94663 CACCACTATTGGCGTTATGATGATGA------CAAACACGGCGATTTATCCTACTCCGGC 94716

Qy    661 gcaagtgcgtattcttggctcgttggtggcaataccttgcacaaaatggatcaggtggt 720
          |||            ||| |  |||| ||||||||   |   ||   | |   |  |  |

Db  94717 GCA-----------TGGTTAATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAAT 94764

Qy    721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggtttttttaccaaca 780
          |||  ||| |  ||  |  | ||        |    | |   |||||   | | || |
Db  94765 GGCGTAGTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATT 94824

Qy    781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
          | |||  |    |||||||| || || ||||||||||| |||||      ||| | ||
Db  94825 GCAGGTGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAA 94884

Qy    841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
          ||| |  | || || || || ||||| ||| |||| |||   || |     || || |||
Db  94885 TGGCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTC 94944

Qy    901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
          |||||| | || |||||||||||||| ||||| || |     || ||||| ||| | ||
Db  94945 CAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACATACCGTC 95004

Qy    961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
          || |  ||||| ||    || || | |  || ||||    | | || || || || ||
Db  95005 TTTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACAACGGTACGGGT 95064

Qy   1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
          |   | |  |  |    | ||| |   |||    || ||      | |||  ||  || ||
Db  95065 ACGGTAACAGAAACCAACGAAAAGGTTTCCAATCCAAA---GCTTAAAGTACAGACAGTC 95121

Qy   1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
          | | ||||| | |  |||||    || || |  | ||||||  ||||   | || |||||
Db  95122 CGACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTT---ACGCGGCAGGG 95178

Qy   1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttccttttattgacgaa 1200
          ||||| ||    || || || ||  | || || || |||||  ||||  ||||| ||  |
Db  95179 GGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGATTAC 95238

Qy   1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
          || || ||| || | ||  |  | | |||||||||| ||||| || || || || |||||
Db  95239 GGCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 95298

Qy   1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
          |||| |||||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db  95299 GAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 95358

Qy   1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95359 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 95418

Qy   1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95419 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 95478

Qy   1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95479 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 95538

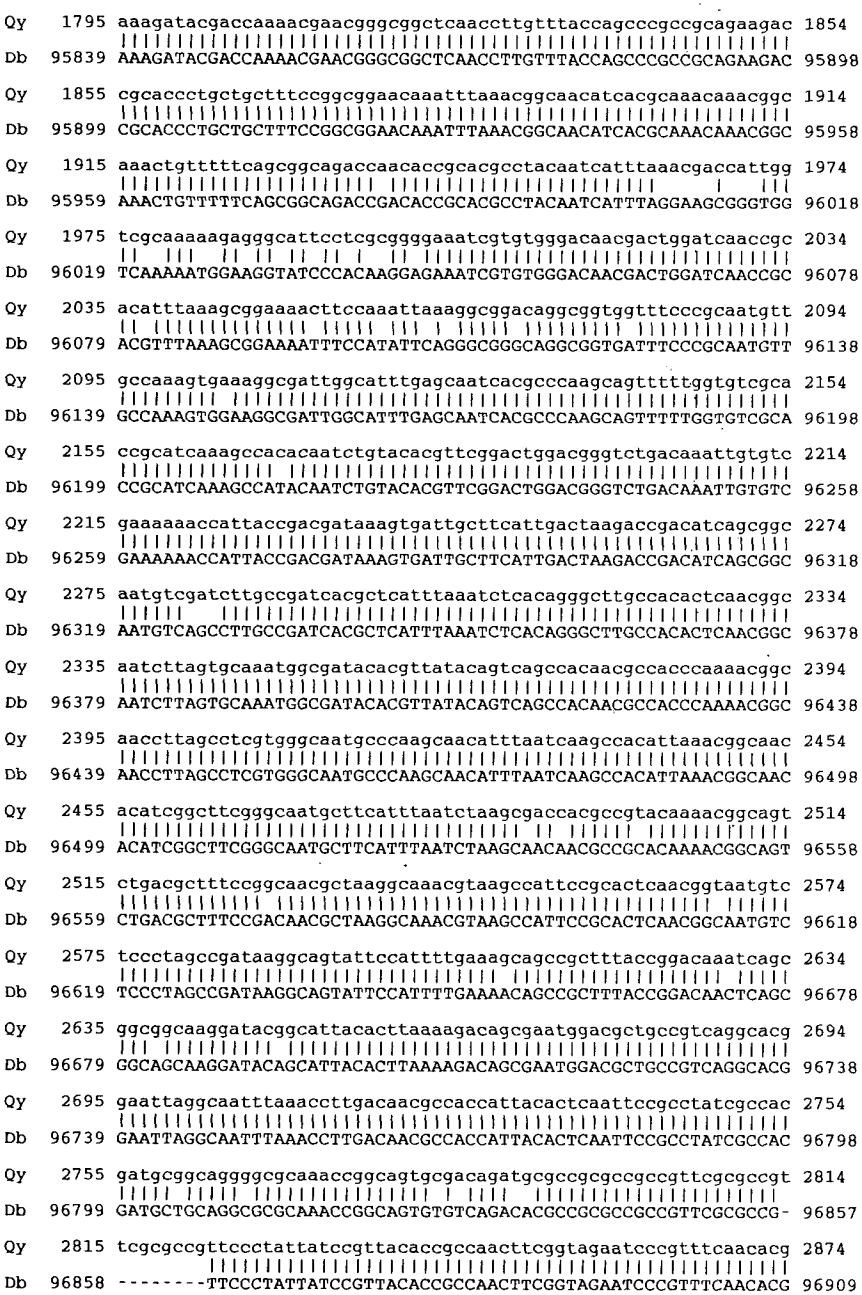
Qy   1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95539 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 95598

Qy   1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95599 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 95658

Qy   1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||| || |||
Db  95659 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAATCATAAT 95718

Qy   1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggc--- 1737
             | |  |||||||||||||||||||| ||| ||  |||| |   | |  |
Db  95719 GCCACAACAACATCCACCGTTACCATTACAGGGAATGAAAGTATTACACAACCGAGTGGT 95778

Qy   1738 ---aataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
             |||| ||| ||  |  ||  ||  |||||||||||||||||||||||||||||||||
Db  95779 AAGAATATCAATAGACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 95838
```


Qy 1795 aaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagac 1854
Db 95839 AAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGAC 95898
Qy 1855 cgcaccctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggc 1914
Db 95899 CGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGC 95958
Qy 1915 aaactgtttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattgg 1974
Db 95959 AAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGG 96018
Qy 1975 tcgcaaaaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgc 2034
Db 96019 TCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCAACCGC 96078
Qy 2035 acatttaaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgtt 2094
Db 96079 ACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTT 96138
Qy 2095 gccaaagtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgca 2154
Db 96139 GCCAAAGTGGAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCA 96198
Qy 2155 ccgcatcaaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtc 2214
Db 96199 CCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 96258
Qy 2215 gaaaaaaccattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggc 2274
Db 96259 GAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGC 96318
Qy 2275 aatgtcgatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggc 2334
Db 96319 AATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 96378
Qy 2335 aatcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggc 2394
Db 96379 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGC 96438
Qy 2395 aaccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2454
Db 96439 AACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 96498
Qy 2455 acatcggcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagt 2514
Db 96499 ACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACAAAACGGCAGT 96558
Qy 2515 ctgacgctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtc 2574
Db 96559 CTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTC 96618
Qy 2575 tccctagccgataaggcagtattccatttgaaagcagccgctttaccggacaaatcagc 2634
Db 96619 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCAGC 96678
Qy 2635 ggcggcaaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacg 2694
Db 96679 GGCAGCAAGGATACAGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACG 96738
Qy 2695 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccac 2754
Db 96739 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 96798
Qy 2755 gatgcggcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgt 2814
Db 96799 GATGCTGCAGGCGCGCAAACCGGCAGTGTGTCAGACACGCCGCGCCGCCGTTCGCGCCG- 96857
Qy 2815 tcgcgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacg 2874
Db 96858 --------TTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 96909
Qy 2875 ctgacggtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttc 2934
Db 96910 CTGACGGTAAACGGCAAATTGAACGGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTC 96969
Qy 2935 ggctaccgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcg 2994
Db 96970 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCG 97029
Qy 2995 gtcaacaataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaa 3054
Db 97030 GTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAA 97089
Qy 3055 gacaacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgcc 3114
Db 97090 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCC 97149
Qy 3115 ggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtc 3174
Db 97150 GGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC 97209
Qy 3175 aaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaa 3234
Db 97210 AAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 97269
Qy 3235 gacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagaca 3294
Db 97270 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACA 97329
Qy 3295 gaaagcgttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcg 3354
Db 97330 GAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 97389
Qy 3355 gaggaagagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaa 3414
Db 97390 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACAGCGCCTTGGCGAAACAGCGCGAA 97449
Qy 3415 gcggaaacccggccggctaccaccgccttcccccgcgcccgccgcgcccgccgggatttg 3474
Db 97450 GCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTG 97509
Qy 3475 ccgcaactgcaaccccaaccg------cagccccaaccgcagcgcgacctgatcagccgt 3528
Db 97510 CCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATCAGCCGT 97569
Qy 3529 tatgccaatagcggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacag 3588
Db 97570 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAG 97629
Qy 3589 gacgaattagaccgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatc 3648
Db 97630 GACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATC 97689
Qy 3649 cgggacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctg 3708
Db 97690 CGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 97749
Qy 3709 cgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcac 3768
Db 97750 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCAC 97809
Qy 3769 aaccggaccgaaaacaccttcgacgacggcatcggcaactcggcacggcttgcccacggc 3828
Db 97810 AACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 97869
Qy 3829 gccgttttcgggcaatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggt 3888
Db 97870 GCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGT 97929
Qy 3889 tttagcagcggcagcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcat 3948
Db 97930 TTTAGCAGCGGCAGTCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCAT 97989
Qy 3949 tacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatc 4008

```
Db  97990 TACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATC 98049

Qy   4009 ggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgcc 4068
Db  98050 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCC 98109

Qy   4069 acccccggccttgcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaa 4128
Db  98110 ACCCCCGGTCTTGCGTTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAA 98169

Qy   4129 ccggcgcaacacatttccatcacgccttatttgagcctgtcctataccgatgccgcttcg 4188
Db  98170 CCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCG 98229

Qy   4189 ggcaaagtccgaacacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgc 4248
Db  98230 GGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGC 98289

Qy   4249 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgcc 4308
Db  98290 AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCC 98349

Qy   4309 gccgccaaaggcccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4368
Db  98350 GCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGC 98409

Qy   4369 tggtaa 4374
Db  98410 TGGTAA 98415


RESULT  10
A96618
LOCUS       A96618                  4350 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 651 from Patent WO9924578.
ACCESSION   A96618
VERSION     A96618.1  GI:6780201
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4350)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 651 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4350
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     1225 a   1161 c   1060 g    842 t     62 others
ORIGIN


  Query Match             80.2%;  Score 3508;  DB 6;  Length 4350;
  Best Local Similarity   88.4%;  Pred. No. 0;
  Matches 3876;  Conservative    0;  Mismatches  462;  Indels   48;  Gaps    8;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
Db     61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
Db    241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
Db    301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy    361 aacgttgattttggtgcggaaggaagaaatcccgatcaacatcgtttacttataaaatt 420
Db    361 AACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAAATT 420

Qy    421 gtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcat 480
Db    421 GTGAAAAGAAATAATTATAAGCCTGACAATTCA---CACCCTTACAACGGCGATTANCAT 477

Qy    481 atgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatg 540
Db    478 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATG 537

Qy    541 gatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggc 600
Db    538 AGGGGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGA 597

Qy    601 aggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatcatatt 660
Db    598 CACCACTATTGGCGTTATGATGATGA------CAAACACGGCGATTTATCCTACTCCGGC 651

Qy    661 gcaagtgcgtattcttggctcgttggtggcaataccttgcacaaaatggatcaggtggt 720
Db    652 GCA------------TGGTTAATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAAT 699

Qy    721 ggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggttttttaccaaca 780
Db    700 GGCGTANTTAGTTTGAGCGGCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATT 759

Qy    781 ggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaag 840
Db    760 GCAGGTGCGGCAGGCGACAGCGGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAA 819

Qy    841 tggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttc 900
Db    820 TGGCTGCTCAACGGAGTTTTACAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTC 879

Qy    901 cagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattcagta 960
Db    880 CAGCTGATACGCAAAGATTGGTTCTACGATGACATTTACAGAGGCGATACACATACCGTC 939

Qy    961 ttctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacagga 1020
Db    940 TNTTTTGAACCGCGCAGTAACGGACATTTTTCCTTTACATCCAACAACAACGGTACGGGT 999

Qy   1021 aaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgtt 1080
Db   1000 ACGGTAACAGAAACCAACGAAAAGGTNTCCAATCCAAA---GCTTAAAGTACAGACAGTC 1056

Qy   1081 caattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggt 1140
Db   1057 CGACTGTTTGACGAATCTTTGAATGAAACTGATAAAGAACCAGTTT---ACGCGGCAGGG 1113

Qy   1141 ggtgtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacgaa 1200
Db   1114 GGTGTTAATCAGTACCGTCCAAGGTTAAACAACGGTGAAAACCTTTCTTTTATCGATTAC 1173

Qy   1201 ggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttc 1260
Db   1174 GGCAACGGCAAACTCATCTTATCAAACAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 1233

Qy   1261 caaggagattttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
```

```
Db    1234 GAAGGTGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1293

Qy    1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1294 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1353

Qy    1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1354 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1413

Qy    1441 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1414 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1473

Qy    1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1474 TTTAGTGAAATCGGCTTGNTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1533

Qy    1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1534 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1593

Qy    1621 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaat 1680
           |||||||||||||||||||||||||||||||||||||||||||||||||| | | || |||
Db    1594 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGNCNATCATAAT 1653

Qy    1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggc--- 1737
           | |  ||||||||||||||||||||||| ||| ||  |||| |   | | |
Db    1654 GCCACAACAACATCCACCGTTACCATTACAGGGAATGAAAGTATTACACAACCGAGTGGT 1713

Qy    1738 ---aataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
              |||| ||| ||  |  ||  ||  ||||||||||||||||||||||||||||||||||
Db    1714 AAGAATATCAATAGACTTAATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1773

Qy    1795 aaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagac 1854
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1774 AAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGAC 1833

Qy    1855 cgcaccctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggc 1914
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1834 CGCACCCNGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGC 1893

Qy    1915 aaactgtttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattgg 1974
           |||||||||||||||||||||| ||||||||||||||||||||||||     |   |||
Db    1894 AAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGG 1953

Qy    1975 tcgcaaaaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgc 2034
           ||  |||  || || ||  ||  |  ||  |||||||||||||||||||||||||||||||||
Db    1954 TCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGACTGGATCNACCGC 2013

Qy    2035 acatttaaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgtt 2094
           || ||||||||||||||| |||||  ||| | ||||| |||||||||||||||||||||||
Db    2014 ACGTTTAAAGCGGAAAATTTCCATATTCAGGGCGGGCAGGCGGTGATTTCCCGCAATGTT 2073

Qy    2095 gccaaagtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgca 2154
           |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    2074 GCCAAAGTGGAAGGCGATTGNCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCA 2133

Qy    2155 ccgcatcaaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtc 2214
           ||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||||||
Db    2134 CCGCATCAAAGCCATACAATCTGTACACGTTCGGACTGGACNGGTCTGACAAATTGTGTC 2193

Qy    2215 gaaaaaaccattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggc 2274
           ||| ||| ||||||||||||||||||||||||||||||||||||||| ||| | ||||||
Db    2194 GAANAAANCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACNGACNTNAGCGGC 2253

Qy    2275 aatgtcgatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggc 2334
           | |||    ||  || || ||| |  || ||| ||| | |||| |||  |||| || |||
Db    2254 ANTGTNAGNCTNNCCNATNACGNTNNTTNAAANCTCNCNGGGCNTGCNNCACTNAANGGC 2313

Qy    2335 aatcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggc 2394
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2314 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGC 2373

Qy    2395 aaccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2454
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2374 AACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 2433

Qy    2455 acatcggcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagt 2514
           |||||| ||||||||||||||||||||||||||| || |||||| ||||||||||||||
Db    2434 NCATCGGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCACAAAACGGCAGT 2493

Qy    2515 ctgacgctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtc 2574
           |||||||||||||| ||||||||||||||||||||||||||||||||||||| ||||||
Db    2494 CTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTC 2553

Qy    2575 tccctagccgataaggcagtattccattttgaaagcagccgctttaccggacaaatcagc 2634
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||| |||||
Db    2554 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGACAACTCAGC 2613

Qy    2635 ggcggcaaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacg 2694
           ||| |||||||  || ||||||||||||||||||||||||||||||||||||||||||||||
Db    2614 GGCAGCAAGGANACAGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACG 2673

Qy    2695 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccac 2754
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2674 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2733

Qy    2755 gatgcggcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgt 2814
           ||||| ||||| ||||||||||||| | | ||||  ||||||||||||||||||||||||
Db    2734 GATGCTGCAGGCGCGCAAACCGGCAGNGTGTCAGACACGCCGCGCCGCCGTTCGCGCCGT 2793

Qy    2815 tcgcgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacg 2874
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2794 ---------TCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 2844

Qy    2875 ctgacggtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttc 2934
           |||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db    2845 CTGACGGTAAACGGCAAATTGAACNGTCAAGGAACATTCCGCTTTATGTCGGAACTCTTC 2904

Qy    2935 ggctaccgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcg 2994
           |||||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    2905 GGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTTGGCG 2964

Qy    2995 gtcaacaataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaa 3054
           |||||||||||||||||||||| | |||||||| |||||||||||||||||||||||| |||
Db    2965 GTCAACAATACCGGCAACGAACCCGTAAGCCTCGATCAATTGACGGTAGTGGAAGGGAAA 3024

Qy    3055 gacaacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgcc 3114
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3025 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCC 3084

Qy    3115 ggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctgcataataccggtc 3174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3085 GGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC 3144

Qy    3175 aaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaa 3234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3145 AAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 3204

Qy    3235 gacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcgaaaagaca 3294
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    3205 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGCCGAAAAGACA 3264

Qy    3295 gaaagcgttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattatgcaggcg 3354
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    3265 GAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3324

Qy    3355 gaggaagagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaacagcgcgaa 3414
           |||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||
Db    3325 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAA 3384

Qy    3415 gcggaaacccggccggctaccaccgccttccccccgcgcccgccgcgcccgccgggatttg 3474
           |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    3385 GCGGAAACCCGGCCGGNTACCACCGCCTTCCCCCGCGCCCGCNGCGCCCGCCGGGATTTG 3444
```

```
Qy   3475 ccgcaactgcaaccccaaccg------cagccccaaccgcagcgcgacctgatcagccgt 3528
          ||||||| ||| |||||||||      || ||||||||||||||||||||| ||||||
Db   3445 CCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAACCGCAGCGCGACCTGATNAGCCGT 3504

Qy   3529 tatgccaatagcggtttgagtgaatttttccgccacgctcaacagcgttttcgccgtacag 3588
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3505 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAG 3564

Qy   3589 gacgaattagaccgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcatc 3648
          |||||||| |||||||| |||||||||||||||||||| |||||||||||| |||||
Db   3565 GACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCNGTTTGGACAAGCNGCATC 3624

Qy   3649 cgggacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctg 3708
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3625 CGGNACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 3684

Qy   3709 cgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcac 3768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3685 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCAC 3744

Qy   3769 aaccggaccgaaaacaccttcgacgacggcatcggcaactcggcacggcttgcccacggc 3828
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3745 AACCGGACCGAAAACANCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3804

Qy   3829 gccgttttcgggcaatacggcatcgacaggttctacatcggcatcagcgcgggcgcgggt 3888
          |||||||||||||||||||||| ||||||| |||||||||||||| ||||||||||||
Db   3805 GCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCACGGGCGCGGGT 3864

Qy   3889 tttagcagcggcagcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgcat 3948
          |||||||||||||  || ||||||||||||||||||||||||||||||||||||||||||
Db   3865 TTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCAT 3924

Qy   3949 tacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgcacatc 4008
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   3925 TACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGTACATC 3984

Qy   4009 ggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgcc 4068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3985 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCC 4044

Qy   4069 acccccggccttgcattcaaccgctaccgcgcgggcattaaggcagattattcattcaaa 4128
          |||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db   4045 ACCCCCGGTCTTGCGTTCAACCGNTACCGNGCGGGCATTAAGGCAGATTATTCATTCAAA 4104

Qy   4129 ccggcgcaacacatttccatcacgccttatttgagcctgtcctataccgatgccgcttcg 4188
          |||||||||||||| |||||||| |||||||| ||||||||||||||||||||||||||||
Db   4105 CCGGCGCAACACATNTCCATCACNCCTTATTTNAGCCTGTCCTATACCGATGCCGCTTCG 4164

Qy   4189 ggcaaagtccgaacacgcgtcaataccgccgtattggctcaggatttcggcaaaacccgc 4248
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   4165 GGCAAAGTCCGAACACGCGTCAATACCGCNGTATTGGCTCAGGATTTCGGCAAAACCCGC 4224

Qy   4249 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctgcc 4308
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   4225 AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCNTCCACGCTGCC 4284

Qy   4309 gccgccaaaggcccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4368
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4285 GCCGCCAAAGGNCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGC 4344

Qy   4369 tggtaa 4374
          ||||||
Db   4345 TGGTAA 4350


RESULT  11
A96614
LOCUS       A96614                 4355 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 647 from Patent WO9924578.
ACCESSION   A96614
VERSION     A96614.1  GI:6780199
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4355)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 647 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4355
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      840 a    775 c    694 g    616 t    1430 others
ORIGIN


  Query Match             61.6%;  Score 2693.2;  DB 6;  Length 4355;
  Best Local Similarity   65.6%;  Pred. No. 0;
  Matches 2916;  Conservative   12;  Mismatches 1353;  Indels  167;  Gaps    6;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGTCGC 60

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagcc 120
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATCCGCTTCTCNGCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy    301 gtggcggcattggtgggcg-atcaatatattgtgagcgtggcacataacggcggctataa 359
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGCGGCATTGGTGGGCGTATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAA 360

Qy    360 caacgttgattttggtgcggaaggaagaaat---cccgatcaacatcgttttacttataa 416
          |||||||||||||||||||||||||||||: |||   || ||||||||:||::|||||||||
Db    361 CAACGTTGATTTTGGTGCGGAAGGAAKNAATATCCCNGATCAACAWCGWWTTACTTATAA 420

Qy    417 aattgtgaaacggaataattataaagcagggactaaaggccatccttatggcggcgatta 476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTA 480

Qy    477 tcatatgccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagtta 536
          ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db    481 TCATATGCCGCGTTTGCATAAATWTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTA 540

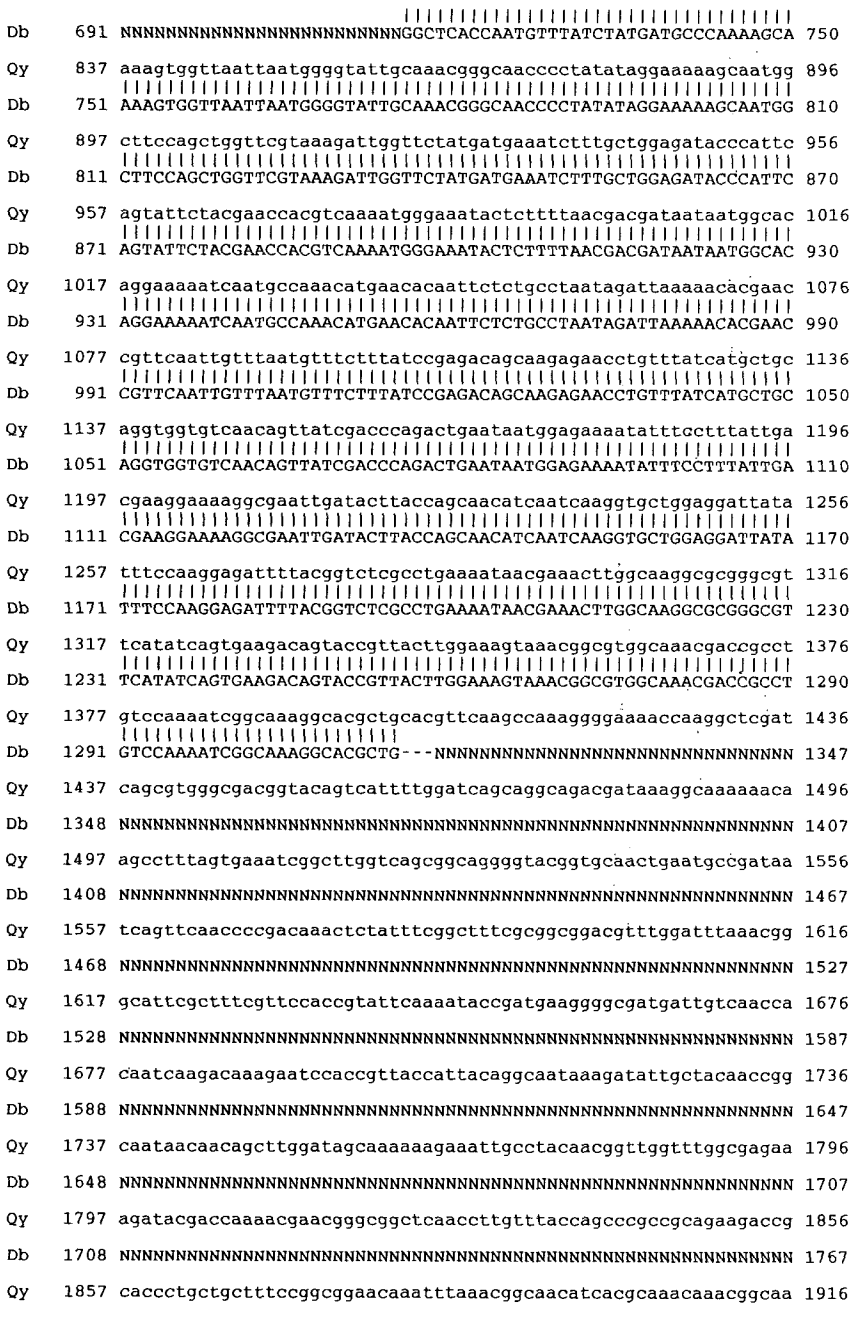
Qy    537 tatggatgggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggc 596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGC 600

Qy    597 aggcaggcaatattggcgatctgatgaagatgagcccaataaccgcgaaagttcatatca 656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCA 660

Qy    657 tattgcaagtgcgtattcttggctcgttggtggcaatacctttgcacaaaatggatcagg 716
          |||||||||
Db    661 TATTGCAAG---------------------------------------------------- 669

Qy    717 tggtggcacagtcaacttaggtagtgaaaaaattaaacatagcccatatggttttttacc 776
                                                  |
Db    670 ----------------------------------------TNNNNNNNNNNNNNNNNNNN 690

Qy    777 aacaggaggctcatttggcgacagtggctcaccaatgtttatctatgatgcccaaaagca 836
```


Db 691 NNNNNNNNNNNNNNNNNNNNNNNNGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 750
Qy 837 aaagtggttaattaatggggtattgcaaacgggcaacccctatataggaaaaagcaatgg 896
Db 751 AAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGG 810
Qy 897 cttccagctggttcgtaaagattggttctatgatgaaatctttgctggagatacccattc 956
Db 811 CTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTC 870
Qy 957 agtattctacgaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcac 1016
Db 871 AGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCAC 930
Qy 1017 aggaaaaatcaatgccaaacatgaacacaattctctgcctaatagattaaaaacacgaac 1076
Db 931 AGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAAC 990
Qy 1077 cgttcaattgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgc 1136
Db 991 CGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGC 1050
Qy 1137 aggtggtgtcaacagttatcgacccagactgaataatggagaaaatatttccttttattga 1196
Db 1051 AGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGA 1110
Qy 1197 cgaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattata 1256
Db 1111 CGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATA 1170
Qy 1257 tttccaaggagatttttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgt 1316
Db 1171 TTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGT 1230
Qy 1317 tcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcct 1376
Db 1231 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCT 1290
Qy 1377 gtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgat 1436
Db 1291 GTCCAAAATCGGCAAAGGCACGCTG---NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1347
Qy 1437 cagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaaca 1496
Db 1348 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1407
Qy 1497 agcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataa 1556
Db 1408 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1467
Qy 1557 tcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgg 1616
Db 1468 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1527
Qy 1617 gcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaacca 1676
Db 1528 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1587
Qy 1677 caatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccgg 1736
Db 1588 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1647
Qy 1737 caataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaa 1796
Db 1648 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1707
Qy 1797 agatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccg 1856
Db 1708 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1767
Qy 1857 caccctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaa 1916
Db 1768 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1827
Qy 1917 actgtttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtc 1976
Db 1828 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1887
Qy 1977 gcaaaaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcac 2036
Db 1888 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1947
Qy 2037 atttaaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgc 2096
Db 1948 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2007
Qy 2097 caaagtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcacc 2156
Db 2008 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2067
Qy 2157 gcatcaaagccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcga 2216
Db 2068 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2127
Qy 2217 aaaaaccattaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaa 2276
Db 2128 NNNNNNNNNNNNNNNNGATAAAGTGACTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2187
Qy 2277 tgtcgatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaa 2336
Db 2188 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAA 2247
Qy 2337 tcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaa 2396
Db 2248 TCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAA 2307
Qy 2397 ccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacac 2456
Db 2308 CCKTAGCCTCGTGGNSAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACAC 2367
Qy 2457 atcggcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtct 2516
Db 2368 ATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCT 2427
Qy 2517 gacgctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctc 2576
Db 2428 GACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTC 2487
Qy 2577 cctagccgataaggcagtattccattttgaaagcagccgctttaccggacaaatcagcgg 2636
Db 2488 CCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGG 2547
Qy 2637 cggcaaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacgga 2696
Db 2548 CGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGARCGGA 2607
Qy 2697 attaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacga 2756
Db 2608 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGA 2667
Qy 2757 tgcggcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttc 2816
Db 2668 TGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2727
Qy 2817 gcgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgct 2876
Db 2728 GCGCCGTTCCCTATTATMCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCT 2787
Qy 2877 gacggtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcgg 2936
Db 2788 GACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGG 2847
Qy 2937 ctaccgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggt 2996
Db 2848 CTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGT 2907

```
Qy   2997 caacaataccggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaaga 3056
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2908 CAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGA 2967

Qy   3057 caacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccgg 3116
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   2968 CAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCAGG 3027

Qy   3117 cgcgtgg---------------------------------------------------- 3123
          |||||||
Db   3028 CGCGTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3087

Qy   3124 ----------------cgttaccaactcatccgcaaagacggcgagttccgcctgcataa 3167

Db   3088 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3147

Qy   3168 tccggtcaaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaacaggc 3227

Db   3148 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3207

Qy   3228 ggaaaaagacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcga 3287

Db   3208 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3267

Qy   3288 aaagacagaaagcgttgccgaaccggcccggcaggcaggcggggaaaatgtcggcattat 3347

Db   3268 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3327

Qy   3348 gcaggcggaggaagagaaaaaacgggtgcaggcggataaagacaccgccttggcgaaaca 3407

Db   3328 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3387

Qy   3408 gcgcgaagcggaaacccggccggctaccaccgccttcccccgcgcccgccgcgcccgccg 3467

Db   3388 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3447

Qy   3468 ggatttgccgcaactgcaaccccaaccgcagccccaaccgcagcgcgacctgatcagccg 3527

Db   3448 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3507

Qy   3528 ttatgccaatagcggtttgagtgaattttccgccacgctcaacagcgttttcgccgtaca 3587

Db   3508 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3567

Qy   3588 ggacgaattagaccgcgtatttgccgaagaccgccgcaacgccgtttggacaagcggcat 3647
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3568 NNNNNNNTTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCAT 3627

Qy   3648 ccgggacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacct 3707
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3628 CCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCT 3687

Qy   3708 gcgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgca 3767
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3688 GCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCA 3747

Qy   3768 caaccggaccgaaaacaccttcgacgacggcatcggcaactcggcacggcttgcccacgg 3827
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3748 CAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGG 3807

Qy   3828 cgccgttttcgggcaatacggcatcgacaggttctacatcggcatcag-cgcgggcgcgg 3886
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   3808 CGCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGNCGCGGGCGCGG 3867

Qy   3887 gttttagcagcggcagcctttcagacggcatcggaggcaaaatccgccgccgcgtgctgc 3946
          ||||||||||||||||||||||||||||||||::|||:|||||||||||||||||||||||
Db   3868 GTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGSMAAAWTCCGCCGCCGCGTGCTGC 3927

Qy   3947 attacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgcaca 4006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3928 ATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACA 3987

Qy   4007 tcggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcg 4066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3988 TCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCG 4047

Qy   4067 ccacccccggccttgcattcaaccgctaccgcgcgggcattaaggcagattattcattca 4126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4048 CCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCA 4107

Qy   4127 aaccggcgcaacacatttccatcacgccttatttgagcctgtcctataccgatgccgctt 4186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4108 AACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTT 4167

Qy   4187 cgggcaaagtccgaacacgcgtcaataccgccgtattggctcaggatttcggcaaaaccc 4246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4168 CGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCC 4227

Qy   4247 gcagtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccctccacgctg 4306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4228 GCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTG 4287

Qy   4307 ccgccgccaaaggcccgcaactggaagcgcaacacagcgcgggcatcaaattaggctacc 4366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4288 CCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACC 4347

Qy   4367 gctggtaa 4374
          ||||||||
Db   4348 GCTGGTAA 4355
```

RESULT 12
U32710/c
LOCUS U32710 13818 bp DNA linear BCT 29-MAY-1998
DEFINITION Haemophilus influenzae Rd section 25 of 163 of the complete genome.
ACCESSION U32710 L42023
VERSION U32710.1 GI:1573200
KEYWORDS .
SOURCE Haemophilus influenzae Rd.
ORGANISM *Haemophilus influenzae* Rd
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 13818)
AUTHORS Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A., Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A., Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A., Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E., Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M., Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L., Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M., Smith,H.O. and Venter,J.C.
TITLE Whole-genome random sequencing and assembly of Haemophilus influenzae Rd
JOURNAL Science 269 (5223), 496-512 (1995)
MEDLINE 95350630
REFERENCE 2 (bases 1 to 13818)
AUTHORS Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodovsky,M., Rudd,K.E. and Koonin,E.V.
TITLE Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli
JOURNAL *Curr. Biol.* 6 (3), 279-291 (1996)
MEDLINE 96398784
REFERENCE 3 (bases 1 to 13818)
AUTHORS White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 4 (bases 1 to 13818)
AUTHORS White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
REMARK The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned, product names have been

```
            improved, and a number of frame shifts have been corrected. We
            gratefully acknowledge the work of Tatusov et. al. We have
            incorporated their annotation into the /notes fields of the
            corresponding H. influenzae genes
REFERENCE   5  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
            Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The whole genome was shifted by 588 nucleotides for a new start
COMMENT     On Sep 30, 1996 this sequence version replaced gi:1220911.
FEATURES             Location/Qualifiers
     source          1. .13818
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            complement(134. .343)
                     /gene="HI0235"
     CDS             complement(134. .343)
                     /gene="HI0235"
                     /note="similar to GB:L29458 SP:P36675 PID:473421 percent
                     identity: 73.81; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21905.1"
                     /db_xref="GI:1573201"
                     /translation="MRKKQKSAVENETVYLHTRGVIKDNAVMALLGDKLFRQRVEKKR
                     KGKGSYQRKAKHPGKIFEKPDYKFF"
     gene            complement(411. .761)
                     /gene="HI0236"
     CDS             complement(411. .761)
                     /gene="HI0236"
                     /note="similar to PID:1805555 percent identity: 45.74;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="arsenate reductase, putative"
                     /protein_id="AAC21906.1"
                     /db_xref="GI:1573202"
                     /translation="MSVIIYHNPHCSKSRETLALLENKGIQPIIELYLQKQYSVNELQ
                     SIAKKLGIDDVRQMMRTKDELYKSLNLDNLDLSQAELFKAISEHSALIERPIVINGDK
                     AKIGRPPETVLEIL"
     gene            840. .1888
                     /gene="HI0237"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     GB:U00096 SP:P77406 PID:1621505 PID:1788838 percent
                     identity: 55.75; identified by sequence similarity;
                     putative"
                     /pseudo
     gene            complement(2003. .2980)
                     /gene="HI0239"
     CDS             complement(2003. .2980)
                     /gene="HI0239"
                     /note="similar to SP:P19674 GB:X56175 PID:581231 GB:U00096
                     PID:1773093 percent identity: 48.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-export membrane protein (secF)"
                     /protein_id="AAC21907.1"
                     /db_xref="GI:1573204"
                     /translation="MMKLFTKDKDGHFIREINGIKLPFPLTEFMKVRKLGYILSALLM
                     VISLFFIITKGFNWGLDFTGGVVFDTHFSQSANLEQIRSKLHENGIESPIVQTTGSVQ
                     DVMIRLPASNNDSTIGEHVKSMLQNVDKDIQIRSIEFVGPNVGEELAQGAVYATLATL
                     AMVLIYVGSRFEWRLGFGSIASLAHDVIITLGVFSALQIEIDLTFVAAILSVVGYSIN
                     DSIVVFDRVRENFRKIRRLDTIDIIDISLTQTLSRTIITSVTTLVVVMALFFFGGPSI
                     HNFSLALLVGIGFGTYSSIFVAIAIAYDVGLRREHMIPPKVDKEIDELP"
     gene            complement(2988. .4838)
                     /gene="HI0240"
     CDS             complement(2988. .4838)
                     /gene="HI0240"
                     /note="similar to SP:P19673 GB:S68715 GB:X56175 PID:581230
                     PID:545175 percent identity: 59.74; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-export membrane protein (secD)"
                     /protein_id="AAC21908.1"
                     /db_xref="GI:1573205"
                     /translation="MLNRYPLWKNLMVIFIVAIGILYSLPNIYGEDPAVQISGTRGQE
                     ANTSVLGQVQDVLKTNNLPTKSIVLENGSILARFTNTDDQLLAKDKIAERLGNNYTTA
                     LNLAPATPAWLSMFGANPMKWGLDLRGGVRFLMEVDMNATLVKRQEQLQDSLRGELRK
                     EKIQYTAIKNTEHFGTLITLANVSQRAKAERIIRQLHPTLDITEPDADSINLGLSTAA
                     LNEARDLAIEQNLTILRKRVAELGVAEAVIQRQGAERIVIELPGVQDTARAKEILGAT
                     ATLEFRIVNQNVTADAISRNMLPADSEVKYDRQGHPVALFKRAVLGGEHIINSSSGLD
                     QHSSTPQVSVTLDSEGGEIMSQTTKKYYKKPMATLYVEYKDNGKKDENGKTILEKHEE
                     VINVATIQGRFGSNFQITGVDSIAEAHNLSTLLKSGALIAPIQIVEERTIGPSLGAQN
                     VEQGINASLWGLVAVIAFMLFYYKMFGVIASFALVINIVLLVGLMSILPGATLSMPGI
                     AGIVLTLGMSVDANVLIFERIKEEIRNGRSIQQAINEGYNGAFTSIFDANLTTILTAI
                     ILYAVGTGPIQGFAITLSLGVAISMFTAITGTRALVNALYGGKQLKKLLI"
     gene            complement(4908. .5201)
                     /gene="HI0241"
     CDS             complement(4908. .5201)
                     /gene="HI0241"
                     /note="similar to PID:599589 percent identity: 62.37;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21909.1"
                     /db_xref="GI:1573206"
                     /translation="MEAQSPMSTLFIFVIFGLIFYFMIYRPQAKRNKEHKKLMSELAK
                     GTEVLTAGGVIGKITKVTEGSDSIVIALNDTTEITINRNYIVSVLPKGSLKSL"
     gene            complement(5309. .5530)
                     /gene="HI0242"
     CDS             complement(5309. .5530)
                     /gene="HI0242"
                     /note="similar to GB:L77117 SP:Q58397 PID:1591652 percent
                     identity: 35.71; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21910.1"
                     /db_xref="GI:1573207"
                     /translation="MKYQLNLTALRCPIPLLSAKKALKNLDKNDELMLILNLESAVEN
                     FSIFAEENSVALVEQYYASEKEFIVILKK"
     gene            complement(5527. .6045)
                     /gene="HI0243"
     CDS             complement(5527. .6045)
                     /gene="HI0243"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. influenzae predicted coding region HI0243"
                     /protein_id="AAC21913.1"
                     /db_xref="GI:1573208"
                     /translation="MQILEPQQFATWNEPIEMLYACHSKVKRFCRQLSILPDYLEKHG
                     YTQAVLNDVEQILSYFNRAAPLHHDDEELDFFPQLVKVAPQTQTSIDELEKQHEYLHE
                     NWNALSVQLEELISEQRQDIDKHLIERFIQGYDRHIALEEPLFEMGRECLSADILTEM
                     GKHMSARRQVKE"
     gene            complement(6110. .7258)
                     /gene="HI0244"
     CDS             complement(6110. .7258)
                     /gene="HI0244"
                     /note="similar to GB:M63939 SP:P19675 GB:S68715 GB:X56175
                     PID:147966 percent identity: 81.60; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tRNA-guanine transglycosylase (tgt)"
                     /protein_id="AAC21911.1"
```

```
                     /db_xref="GI:1573209"
                     /translation="MKYELDKTSGNARRGRLVFERPQGTFSVETPAFMPVGTYGTVKG
                     MTPEEVRATGAEILLGNTFHLWLRPGQEVMRKHGDLHDFMQWHRPILTDSGGFQVFSL
                     GKLRKITEEGVKFQNPINGERIFLSPEKSMEIQYDLGSDIVMIFDECTPYPATFDYAK
                     KSMEMSLRWAKRSRDRFDELGNKNALFGIIQGGVFEELRKVSLEGLVNIGFDGYAVGG
                     LAVGEPKEDMHRILEYICPQIPADKPRYLMGVGKPEDLVEGVRRGIDMFDCVMPTRNA
                     RNGHLFVTDGIVKIRNAKYRDDTSPLDPECDCYTCKNYTKAYLYHLDKCGEILGARLN
                     TIHNLRYYQRLMAEIRQAIEDDRFDDFVVEFYARMGKPVPPLQLADKS"
     gene            complement(7602. .8693)
                     /gene="HI0245"
     CDS             complement(7602. .8693)
                     /gene="HI0245"

  Query Match             29.1%;  Score 1274.4;  DB 1;  Length 13818;
  Best Local Similarity   58.9%;  Pred. No. 2.2e-276;
  Matches 2572;  Conservative    0;  Mismatches 1571;  Indels  225;  Gaps   13;

Qy      73 cctgcttacttagccatatgcctgtcgttcggcattcttccccaagcctgggcgggacac 132
           | |  ||  ||| ||   ||  | || || || ||     | ||||||||||| || ||
Db   13610 CTTAATTTTTTAACCGCTTGTGTTTCATTAGGGATAGCATCACAAGCCTGGGCAGGTCAT 13551

Qy     133 acttatttcggcatcaactaccaatactatcgcgactttgccgaaaataaaggcaagttt 192
           |||||||| || ||  |||||||||| ||||| || |||||||| |||||||| |||||
Db   13550 ACTTATTTTGGGATTGACTACCAATATTATCGTGATTTTGCCGAGAATAAAGGGAAGTTC 13491

Qy     193 gcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatca 252
           |||| ||||| ||| |||||||||||||| |||||| ||||| | || || |||| ||||
Db   13490 ACAGTTGGGGCTAAAAATATTGAGGTTTATAACAAAGAAGGGCAATTAGTTGGCACATCA 13431

Qy     253 atgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattg 312
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   13430 ATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCATTA 13371

Qy     313 gtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgatttt 372
           || |||||||| |||||||||||||||||||||||||||||||| |||||| | |||||||||
Db   13370 GTAGGCGATCAGTATATTGTGAGCGTGGCACATAACGGCGGATATAACGATGTTGATTTT 13311

Qy     373 ggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaattgtgaaacggaat 432
           ||||| |||||| |||| || ||||| || || |||||||||| ||||||| ||| | |||
Db   13310 GGTGCAGAAGGACGAAACCCTGATCAGCACCGCTTTACTTATCAAATTGTAAAAAGAAAT 13251

Qy     433 aattataaagcagggactaaaggccatccttatggcggcgattatcatatgccgcgtttg 492
           |||||| |||| ||  | |  |||||||||| || ||||||||||||| |||||
Db   13250 AATTATCAAGCTTGGGAGAGAAAGCATCCTTATGATGGAGATTATCATATGCCTCGTTTA 13191

Qy     493 cataaatttgtcacagatgcagaacctgttgaaatgaccagttatatggatgggcggaaa 552
           |||||||||||| || || || |||||||| |  ||||| |   ||||||||||      |
Db   13190 CATAAATTTGTAACTGAAGCTGAACCTGTGGGTATGACAACAAATATGGATGGAAAAGTA 13131

Qy     553 tatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggcaggcaatattgg 612
           |||  |||  | | || || ||||| ||||| ||||| ||  ||||  | || ||||||
Db   13130 TATGCTGATAGAGAGAACTATCCTGAGCGTGTACGTATAGGCTCAGGACGTCAGTATTGG 13071

Qy     613 cgatctgatgaagatgagcccaataaccgcgaaagttcatatcatattgcaagtgcgtat 672
           ||  | ||| |||||||    |  ||     | |||||||||| || |  || |||| |||
Db   13070 CGTACAGATAAAGATGAAGAAACGAATGTACATAGTTCATATTATGTCTCAGGTGCATAT 13011

Qy     673 tcttggctcgttggtggcaatacctttgcacaaaatggatcaggtggtggcacagtcaac 732
             ||  ||   ||  || ||||||  | | || | ||||   |||  ||| |||||||||
Db   13010 CGTTATCTTACTGCAGGAAATACCCATACTCAGAGTGGAAATGGTAATGGTACAGTCAAT 12951

Qy     733 ttaggtagtgaaaaaattaaacatagcccatatggtttttttaccaacaggaggctcattt 792
           |  || || |   | |||  | ||  |  ||||||   |||||||| || || ||
Db   12950 CTTAGTGGTAATGTAGTTAGCCCTAATCATTATGGTCCATTACCAACGGGTGGTTCTAAA 12891

Qy     793 ggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaagtggttaattaat 852
           ||||| || || || |||||||||||| ||||||||  | ||| || | ||| | || |||
Db   12890 GGCGATAGCGGTTCGCCAATGTTTATTTATGATGCGAAGAAGAAACAATGGCTTATAAAT 12831

Qy     853 ggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttccagctggttcgt 912
           |  ||||| ||||| ||  | || | | | |||| | | ||||| || ||| |  | |||
Db   12830 GCTGTATTACAAACTGGGCATCCTTTTTTCGGAAGAGGTAATGGGTTTCAGTTAATACGT 12771

Qy     913 aaagattggttctatgatgaaatctttgctggagatacccattcagtatt------ctac 966
           |||| ||||| ||| ||||| |  |||| |  |||||||| |   || ||      |||
Db   12770 GAAGAATGGTTTTATAATGAAGTTCTTGCGGTTGATACCCCTAGTGTTTTTCAACGCTAT 12711

Qy     967 gaaccacgtcaaaatgggaaatactcttttaacgacgataataatggcacaggaaaaatc 1026
             || |    |||||| | || || |||       ||||| |||| ||||| ||| |
Db   12710 ATTCCCCCAATAAATGGACATTATTCCTTTGTATCAAATAATGATGGTACAGGTAAATTA 12651

Qy    1027 aat---gccaaacatgaacacaattctctgcctaatagattaaaaacacgaaccgttcaa 1083
           | |   | | || |    |  ||            | || | ||  | |||| ||  |
Db   12650 ACTTTAACTAGACCTAGTAAAGATGGCTCTAAAGCAAAATCAGAAGTAGGAACTGTGAAG 12591

Qy    1084 ttgtttaatgtttctttatccgagacagcaagagaacctgtttatcatgctgcaggtggt 1143
           || |||||   || |||    | || || | ||||| |||   | | || |||| ||
Db   12590 TTATTTAATCCATCGTTAAATCAAACCGCTAAAGAACATGT---TAAAGCAGCAGCAGGC 12534

Qy    1144 gtcaacagttatcgacccagactgaataatggagaaaatatttcctttattgacgaagga 1203
             || | ||| | ||| ||| || |  ||||| |||||||||| | ||   ||| |||||
Db   12533 TATAATATTTACCAACCAAGAATGGAATATGGAAAAAATATTTACCTTGGCGACCAAGGA 12474

Qy    1204 aaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttccaa 1263
           |||||    || | | |    | || || |||||||||||||| |||||||| ||  ||
Db   12473 AAAGGAACTTTAACAATCGAAAATAATATAAATCAAGGTGCTGGTGGATTATACTTTGAA 12414

Qy    1264 ggagattttacggt---ctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcat 1320
           ||  |||||   ||          |||||||  ||||||||||| || |||||   |
Db   12413 GGTAATTTTGTTGTAAAAGGCAAGCAAAATAATATAACTTGGCAAGGTGCAGGCGTATCT 12354

Qy    1321 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1380
           ||  |  ||||    || |||   |||||||  ||      | ||| || ||  | ||
Db   12353 ATTGGACAAGATGCAACTGTTGAATGGAAAGTTCACAATCCTGAAAATGATCGTTTATCT 12294

Qy    1381 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1440
           ||||| || | ||||||  |    ||  | |  || ||  | ||   |||     | ||
Db   12293 AAAATTGGTATAGGCACTTTATTAGTCAATGGTAAGGGAAAGAATTTAGGAAGTTTAAGT 12234

Qy    1441 gtgggcgacggtacagtcatttggatcagcaggcagacgataaaggcaaaaaacaagcc 1500
           | |||  |||| | ||||||| | ||||| || ||||| ||    ||  ||||||||||
Db   12233 GCGGGTAACGGCAAAGTCATTCTAGATCAACAAGCAGATGAAGCGGGTCAAAAACAAGCT 12174

Qy    1501 tttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcag 1560
           || |  ||| | ||| | || |||||  | |  || || ||| | |||       ||||
Db   12173 TTCAAAGAAGTTGGCATTGTAAGCGGTCGAGCAACAGTTCAATTAAATAGTACAGATCAA 12114

Qy    1561 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1620
           |  | ||  ||||  |||||||||| ||||| || || || || ||| | |||||||||
Db   12113 GTTGATCCTAACAATATCTATTTCGGATTTCGTGGTGGTCGCTTAGATCTTAACGGGCAT 12054

Qy    1621 tcgctttcgttccaccgtattcaaaatacccgatgaaggggcgatgattgtcaaccacaat 1680
           ||  |  | ||  | ||||| |||||||| || || || |||||||||| ||||| |||
Db   12053 TCATTAACCTTTAAGCGTATCCAAAATACGGACGAGGGCGCGATGATTGTGAACCATAAT 11994

Qy    1681 caagacaaagaatccaccgttaccattacaggcaataaagatattgctacaaccggcaat 1740
           |    |||   | |   |||| ||||| || ||  |||| || |  |  |    |||
Db   11993 ACAACTCAAGTCGCTAATATTACTATTACTGGGAACGAAAGTATTACTGCTCCATCTAAT 11934

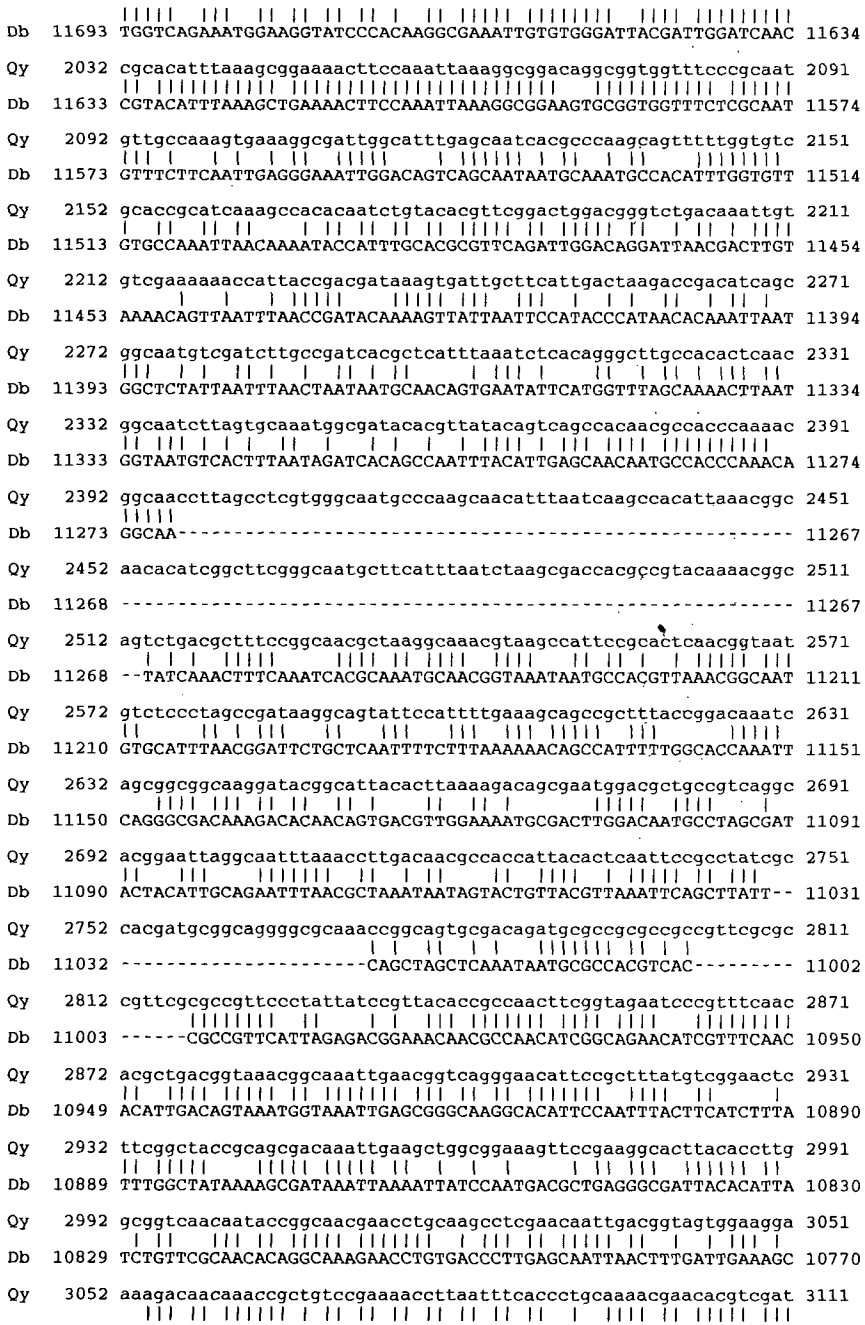
Qy    1741 aacaacagcttggatag---------caaaaaagaaattgcctacaacggttggtttggc 1791
           || || |   | |||          ||  ||||||||||||||||||||||||||||||
Db   11933 AAAAAGAATATTAATAAACTTGATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGC 11874

Qy    1792 gagaaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaa 1851
           || | |||||  |  |||   || || ||  | ||||||  ||||  | ||  || |||||
Db   11873 GAAACAGATAAAAATAAACACAATGGACGATTAAACCTTATTTATAAACCAACCACAGAA 11814

Qy    1852 gaccgcaccctgctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaac 1911
           || || ||  |||| ||||| || || ||||| ||||| ||| | || || |||||||||
Db   11813 GATCGTACTTTGCTACTTTCAGGTGGCACAAACTTAAAAGGCGATATTACCCAAACAAAA 11754

Qy    1912 ggcaaactgtttttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccat 1971
           || ||||| |||||||||||| ||||| ||||| ||||||||||||||||||| || | | |
Db   11753 GGTAAACTATTTTTCAGCGGTAGACCGACACCCCACGCCTACAATCATTTAGACAAACGT 11694

Qy    1972 tggtcgcaaaaagagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaac 2031
```


Db 11693 TGGTCAGAAATGGAAGGTATCCCACAAGGCGAAATTGTGTGGGATTACGATTGGATCAAC 11634
Qy 2032 cgcacatttaaagcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaat 2091
Db 11633 CGTACATTTAAAGCTGAAAACTTCCAAATTAAAGGCGGAAGTGCGGTGGTTTCTCGCAAT 11574
Qy 2092 gttgccaaagtgaaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtc 2151
Db 11573 GTTTCTTCAATTGAGGGAAATTGGACAGTCAGCAATAATGCAAATGCCACATTTGGTGTT 11514
Qy 2152 gcaccgcatcaaagccacacaatctgtacacgttcggactggacgggtctgacaaattgt 2211
Db 11513 GTGCCAAATTAACAAAATACCATTTGCACGCGTTCAGATTGGACAGGATTAACGACTTGT 11454
Qy 2212 gtcgaaaaaaccattaccgacgataaagtgattgcttcattgactaagaccgacatcagc 2271
Db 11453 AAAACAGTTAATTTAACCGATACAAAAGTTATTAATTCCATACCCATAACACAAATTAAT 11394
Qy 2272 ggcaatgtcgatcttgccgatcacgctcatttaaatctcacagggcttgccacactcaac 2331
Db 11393 GGCTCTATTAATTTAACTAATAATGCAACAGTGAATATTCATGGTTTAGCAAAACTTAAT 11334
Qy 2332 ggcaatcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaac 2391
Db 11333 GGTAATGTCACTTTAATAGATCACAGCCAATTTACATTGAGCAACAATGCCACCCAAACA 11274
Qy 2392 ggcaaccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggc 2451
Db 11273 GGCAA---------- 11267
Qy 2452 aacacatcggcttcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggc 2511
Db 11268 ---------- 11267
Qy 2512 agtctgacgctttccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaat 2571
Db 11268 --TATCAAACTTTCAAATCACGCAAATGCAACGGTAAATAATGCCACGTTAAACGGCAAT 11211
Qy 2572 gtctccctagccgataaggcagtattccattttgaaagcagccgctttaccggacaaatc 2631
Db 11210 GTGCATTTAACGGATTCTGCTCAATTTTCTTTAAAAAACAGCCATTTTTGGCACCAAATT 11151
Qy 2632 agcggcggcaaggatacggcattacacttaaaagacagcgaatggacgctgccgtcaggc 2691
Db 11150 CAGGGCGACAAAGACACAACAGTGACGTTGGAAAATGCGACTTGGACAATGCCTAGCGAT 11091
Qy 2692 acggaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgc 2751
Db 11090 ACTACATTGCAGAATTTAACGCTAAATAATAGTACTGTTACGTTAAATTCAGCTTATT-- 11031
Qy 2752 cacgatgcggcaggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgc 2811
Db 11032 ----------CAGCTAGCTCAAATAATGCGCCACGTCAC--------- 11002
Qy 2812 cgttcgcgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaac 2871
Db 11003 ------CGCCGTTCATTAGAGACGGAAACAACGCCAACATCGGCAGAACATCGTTTCAAC 10950
Qy 2872 acgctgacggtaaacggcaaattgaacggtcagggaacattccgctttatgtcggaactc 2931
Db 10949 ACATTGACAGTAAATGGTAAATTGAGCGGGCAAGGCACATTCCAATTTACTTCATCTTTA 10890
Qy 2932 ttcggctaccgcagcgacaaattgaagctggcggaaagttccgaaggcacttacaccttg 2991
Db 10889 TTTGGCTATAAAAGCGATAAATTAAAATTATCCAATGACGCTGAGGGCGATTACACATTA 10830
Qy 2992 gcggtcaacaatacccggcaacgaacctgcaagcctcgaacaattgacggtagtggaagga 3051
Db 10829 TCTGTTCGCAACACAGGCAAAGAACCTGTGACCCTTGAGCAATTAACTTTGATTGAAAGC 10770
Qy 3052 aaagacaacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgat 3111
Db 10769 TTAGATAATAAACCGTTATCAGATAAGCTCAAATTTACTTTAGAAAATGACCACGTTGAT 10710
Qy 3112 gccggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccg 3171
Db 10709 GCAGGTGCATTACGTTATAAATTAGTGAAGAATAAGGGCGAATTCCGCTTGCATAACCCA 10650
Qy 3172 gtcaaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaaacaggcggaa 3231
Db 10649 ATAAAAGAGCAGGAATTGCTCAATGATTTAGTAAGAGCAGAGCAAGCAGAACAAACATTA 10590
Qy 3232 aaag----acaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgtcga 3287
Db 10589 GAAGCCAAACAAGTTGAACAGACTGCTGAAAAACAAAAAAGTAAGGCAAAAGCGCGGTCA 10530
Qy 3288 aaagacagaaagcgttgccgaaccggcccggcaggcaggcgggaa-------aatgtcg 3340
Db 10529 AGAAGAGCGGTGTTGTCTGATACCCCGTCTGCTCAAAGCCTGTTAAACGCATTAGAAGCC 10470
Qy 3341 gcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaagacaccgccttgg 3400
Db 10469 AAACAAGTTGAACAGACTACTGAAACACAAACAAGTAAGCCAAAAACAAAAAAAGGGCGG 10410
Qy 3401 cgaaacagcgcgaag----------cggaaacccggccggctaccaccgccttcccc 3447
Db 10409 TCAAAAAGAGCATTGAGTGCAGCGTTTTCTGATACCCCGTTTGATCTAAGCCAGTTAAAG 10350
Qy 3448 cgcgcccgccgcgcccgccgggattt---------gccgcaactgcaaccccaaccgcag 3498
Db 10349 GTATTCGAAGTCAAACTTGAGGTTATTAATGCCCAACCGCAAGTGAAAAAAGAACCTCAA 10290
Qy 3499 ccccaaccg----------cagcgcgacctgatcagccgttatgccaatagcggtttg 3546
Db 10289 GATCAAGAGGAACAAGGCAAACAAAAAGAATTGATCAGCCGTTACTCAAATAGTGCGTTA 10230
Qy 3547 agtgaattttccgccacgctcaacagcgttttcgccgtacaggacgaattagaccgcgta 3606
Db 10229 TCGGAGTTGTCTGCAACAGTAAATAGTATGTTTTCCGTTCAAGATGAATTGGATCGTCTT 10170
Qy 3607 tttgccgaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaacactac 3666
Db 10169 TTTGTAGATCAAGCACAATCTGCCCTGTGGACAAATATCGCACAGGATAAAAGACGCTAT 10110
Qy 3667 cgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcag 3726
Db 10109 GATTCTGATGCGTTCCGTGCTTATCAGCAGAAAACGAACTTGCGTCAAATTGGGGTGCAA 10050
Qy 3727 aaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaaaacacc 3786
Db 10049 AAAGCCTTAGATAATGGACGAATTGGGGCGGTTTTCTCGCATAGCCGTTCAGATAATACC 9990
Qy 3787 ttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcgggcaatac 3846
Db 9989 TTTGACGAACAGGTTAAAAATCACGCGACATTAACGATGATGTCGGGTTTTGCCCAATAT 9930
Qy 3847 ggcatcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggcagcctt 3906
Db 9929 CAATGGGGCGATTTACAATTTGGTGTAAATGTGGGTGCGGGAATTAGTGCGAGTAAAATG 9870
Qy 3907 tcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcaggcacga 3966
Db 9869 GCTGAAGAACAAAGCCGAAAAATTCATCGAAAAGCGATAAATTATGGAGTGAATGCAAGT 9810
Qy 3967 taccgcgccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttc 4026
Db 9809 TATCAGTTCCGTTTAGGGCAATTGGGTATTCAGCCTTATTTGGGTGTTAATCGATATTTT 9750
Qy 4027 gtccaaaaagcggattaccgctacgaaaacgtcaatatcgccaccccggccttgcattc 4086
Db 9749 ATTGAACGTGAAAATTATCAATCTGAAGAAGTGAAAGTGCAAACACCGAGCCTTGCATTT 9690
Qy 4087 aaccgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacacatttcc 4146
Db 9689 AATCGCTATAATGCTGGCATTCGAGTTGATTATACATTTACCCCGACAAATAATATCAGC 9630

```
Qy   4147 atcacgccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccgaacacgc 4206
          | | ||||||||||     |   |||  ||||   |    |  || || | |||
Db   9629 GTTAAGCCTTATTTCTTTGTCAATTATGTTGATGTTTCAAACGCTAACGTACAAACCACT 9570

Qy   4207 gtcaataccgccgtattggctcaggatttcggcaaaacccgcagtgcggaatggggcgta 4266
          || |||| | |  | |||   ||    || ||        |     |||  |||  ||
Db   9569 GTAAATAGCACGATGTTGCAACAATCATTTGGGCGTTATTGGCAAAAAGAAGTGGGATTA 9510

Qy   4267 aacgccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggcccgcaa 4326
          || || |||||   |  ||||  || ||       | |     |  ||||  |||||
Db   9509 AAGGCAGAAATTTTACATTTCCAACTTTCTGCTTTTATTTCTAAATCTCAAGGTTCGCAA 9450

Qy   4327 ctggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
          || |     || || |  | |||| | ||||| || || || ||||||
Db   9449 CTCGGCAAACAGCAAAATGTGGGCGTGAAATTGGGGTATCGTTGGTAA 9402


RESULT  13
HIU11024
LOCUS       HIU11024                4317 bp    DNA     linear   BCT 12-JUL-1996
DEFINITION  Haemophilus influenzae N187 adhesion and penetration protein (hap)
            gene, complete cds.
ACCESSION   U11024
VERSION     U11024.1  GI:1408568
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 4317)
  AUTHORS   St. Geme,J.W. III, de la Morena,M.L. and Falkow,S.
  TITLE     A Haemophilus influenzae IgA protease-like protein promotes
            intimate interaction with human epithelial cells
  JOURNAL   Mol. Microbiol. 14 (2), 217-233 (1994)
  MEDLINE   95131744
REFERENCE   2  (bases 1 to 4317)
  AUTHORS   St. Geme,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1994) Joseph W. St. Geme, Molecular Microbiology,
            Washington University School of Medicine, 660 South Euclid Ave.,
            St. Louis, MO 63110, USA
COMMENT     On Jul 12, 1996 this sequence version replaced gi:562013.
FEATURES             Location/Qualifiers
     source          1. .4317
                     /organism="Haemophilus influenzae"
                     /strain="N187"
                     /db_xref="taxon:727"
     gene            61. .4242
                     /gene="hap"
     CDS             61. .4242
                     /gene="hap"
                     /standard_name="Haemophilus adhesion and penetration"
                     /note="HAP"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion and penetration protein"
                     /protein_id="AAB03707.1"
                     /db_xref="GI:1408569"
                     /translation="MKKTVFRLNFLTACISLGIVSQAWAGHTYFGIDYQYYRDFAENK
                     GKFTVGAQNIKVYNKQGQLVGTSMTKAPMIDFSVVSRNGVAALVENQYIVSVAHNVGY
                     TDVDFGAEGNNPDQHRFTYKIVKRNNYKKDNLHPYEDDYHNPRLHKFVTEAAPIDMTS
                     NMNGSTYSDRTKYPERVRIGSGRQFWRNDQDKGDQVAGAYHYLTAGNTHNQRGAGNGY
                     SYLGGDVRKAGEYGPLPIAGSKGDSGSPMFIYDAEKQKWLINGILREGNPFEGKENGF
                     QLVRKSYFDEIFERDLHTSLYTRAGNGVYTISGNDNGQGSITQKSGIPSEIKITLANM
                     SLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDQKQGSLIFASDINQGAGGLYFEG
                     NFTVSPNSNQTWQGAGIHVSENSTVTWKVNGVEHDRLSKIGKGTLHVQAKGENKGSIS
                     VGDGKVILEQQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKFYFGFRGGRLDLN
                     GHSLTFKRIQNTDEGAMIVNHNTTQAANVTITGNESIVLPNGNNINKLDYRKEIAYNG
                     WFGETDKNKHNGRLNLIYKPTTEDRTLLLSGGTNLKGDITQTKGKLFFSGRPTPHAYN
                     HLNKRWSEMEGIPQGEIVWDHDWINRTFKAENFQIKGGSAVVSRNVSSIEGNWTVSNN
                     ANATFGVVPNQQNTICTRSDWTGLTTCQKVDLTDTKVINSIPKTQINGSINLTDNATA
                     NVKGLAKLNGNVTLTNHSQFTLSNNATQIGNIRLSDNSTATVDNANLNGNVHLTDSAQ
                     FSLKNSHFSHQIQGDKGTTVTLENATWTMPSDTTLQNLTLNNSTITLNSAYSASSNNT
                     PRRRSLETETTPTSAEHRFNTLTVNGKLSGQGTFQFTSSLFGYKSDKLKLSNDAEGDY
                     ILSVRNTGKEPETLEQLTLVESKDNQPLSDKLKFTLENDHVDAGALRYKLVKNDGEFR
                     LHNPIKEQELHNDLVRAEQAERTLEAKQVEPTAKTQTGEPKVRSRRAARAAFPDTLPD
                     QSLLNALEAKQAELTAETQKSKAKTKKVRSKRAVFSDPLLDQSLFALEAALEVIDAPQ
                     QSEKDRLAQEEAEKQRKQKDLISRYSNSALSELSATVNSMLSVQDELDRLFVDQAQSA
                     VWTNIAQDKRRYDSDAFRAYQQKTNLRQIGVQKALANGRIGAVFSHSRSDNTFDEQVK
                     NHATLTMMSGFAQYQWGDLQFGVNVGTGISASKMAEEQSRKIHRKAINYGVNASYQFR
                     LGQLGIQPYFGVNRYFIERENYQSEEVRVKTPSLAFNRYNAGIRVDYTFTPTDNISVK
                     PYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKSQGSQ
                     LGKQQNVGVKLGYRW"
BASE COUNT     1496 a    776 c    890 g   1155 t
ORIGIN


  Query Match             28.8%;  Score 1260.8;  DB 1;  Length 4317;
  Best Local Similarity   58.8%;  Pred. No. 2.2e-273;
  Matches 2564;  Conservative    0;  Mismatches 1537;  Indels  261;  Gaps   12;

Qy      73 cctgcttacttagccatatgcctgtcgttcggcattcttccccaagcctgggcgggacac 132
           | |  ||  ||| ||   ||| | || || || ||  |  | ||||| ||||| || |||
Db      82 CTTAATTTTTTAACCGCTTGCATTTCATTAGGGATAGTATCGCAAGCGTGGGCTGGTCAC 141

Qy     133 acttatttcggcatcaactaccaatactatcgcgactttgccgaaaataaaggcaagttt 192
           |||||||| || ||  | |||||||| ||||| || |||||||| |||||||| |||||
Db     142 ACTTATTTTGGGATTGATTACCAATATTATCGTGATTTTGCCGAGAATAAAGGGAAGTTC 201

Qy     193 gcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatca 252
           |||| |||||  || ||||| ||||||| |||||| ||||| | || || |||| ||||
Db     202 ACAGTTGGGGCTCAAAATATTAAGGTTTATAACAAACAAGGGCAATTAGTTGGCACATCA 261

Qy     253 atgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattg 312
           |||||||||||||||||||||||||||||||| ||||| |||||||||||||| || |||
Db     262 ATGACAAAAGCCCCGATGATTGATTTTTCTGTAGTGTCACGTAACGGCGTGGCAGCCTTG 321

Qy     313 gtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgatttt 372
           || |  ||||||||||||||||||||||||||||||  || ||||   | |||||||||
Db     322 GTTGAAAATCAATATATTGTGAGCGTGGCACATAACGTAGGATATACAGATGTTGATTTT 381

Qy     373 ggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaattgtgaaacggaat 432
           ||||| || |||| || ||||||||||||||||||||||||||| ||||| ||||| |||
Db     382 GGTGCAGAGGGAAACAACCCCGATCAACATCGTTTTACTTATAAGATTGTAAAACGAAAT 441

Qy     433 aattataaagcagggactaaaggccatccttatggcggcgattatcatatgccgcgtttg 492
           || || |||  ||    |||    |||||||||| | |||||| |||| || || ||
Db     442 AACTACAAAAAAG---ATAATTTACATCCTTATGAGGACGATTACCATAATCCACGATTA 498

Qy     493 cataaatttgtcacagatgcagaacctgttgaaatgaccagttatatggatgggcggaaa 552
           |||||||| || ||||| || |  ||  |||| |||||     ||||| |||| | |
Db     499 CATAAATTCGTTACAGAAGCGGCTCCAATTGATATGACTTCGAATATGAATGGCAGTACT 558

Qy     553 tatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggcaggcaatattgg 612
           |||   |||  ||  || || || || |||||||||| ||  | ||  |||| | ||||
Db     559 TATTCAGATAGAACAAAATATCCAGAACGTGTTCGTATCGGCTCTGGACGGCAGTTTTGG 618

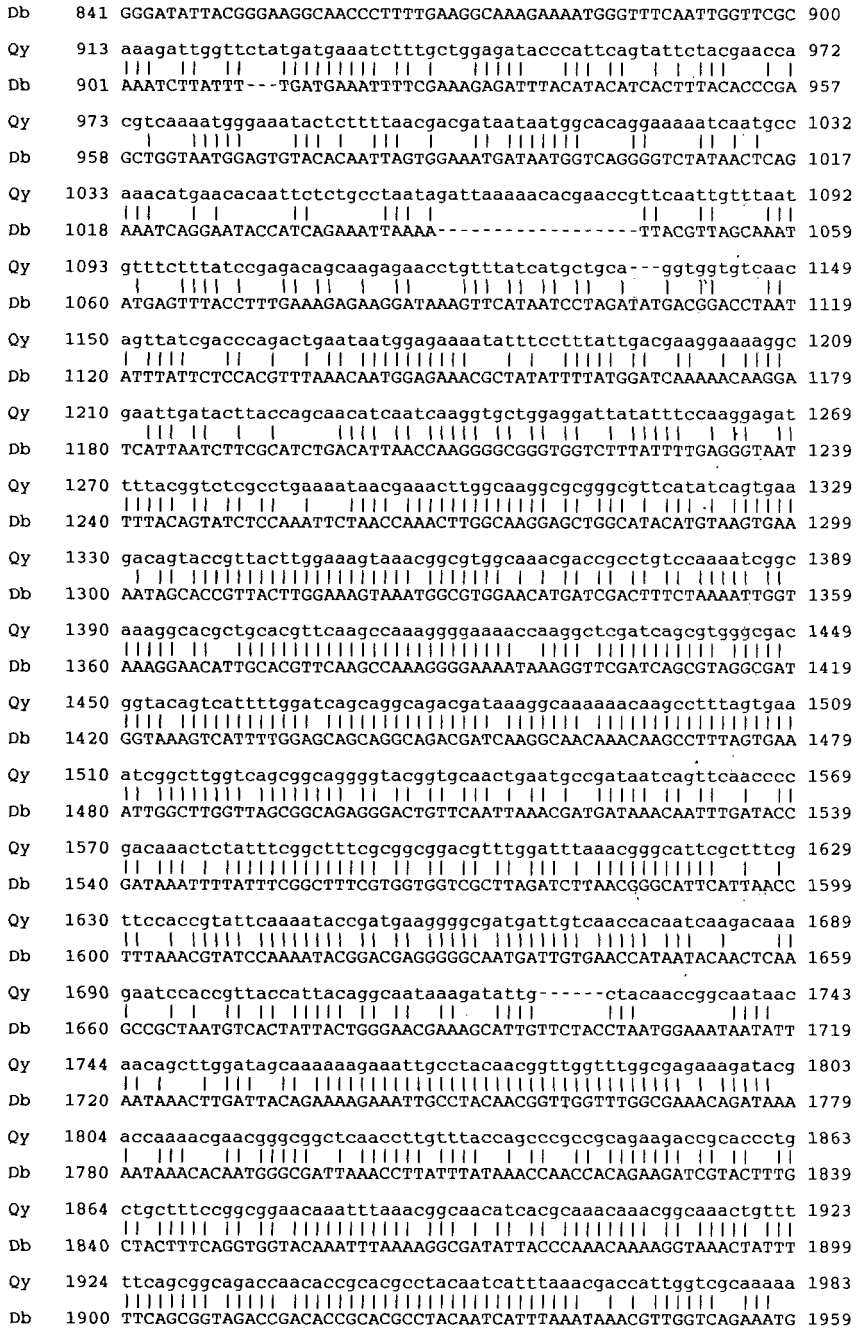
Qy     613 cgatctgatgaagatgagcccaataaccgcgaaagttcatatcatattgcaagtgcgtat 672
           |||  |||| |||| |  |                    | ||  |||| |||| |||
Db     619 CGAAATGATCAAGACAAAGGC----------------GACCAAGTTGCCGGTGCATAT 660

Qy     673 tcttggctcgttggtggcaataccttgcacaaaatggatcaggtggtggcacagtcaac 732
             ||  ||   | |||||||||       ||   |||| |||||  |||        |
Db     661 CATTATCTGACAGCTGGCAATACACACAATCAGCGTGGAGCAGGTAATGGATATTCGTAT 720

Qy     733 ttaggtagtgaaaaaattaaacatagcccatatggtttttaccaacaggaggctcattt 792
           || ||  | ||      ||||    |   |||||||  ||||| |  | |||||||
Db     721 TTGGGAGGCGATGTTCGTAAAGCGGGAGAATATGGTCCATTACCGATTGCAGGCTCAAAG 780

Qy     793 ggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaagtggttaattaat 852
           || |||||||| || || |||||||| |||||||  |||| ||||| ||||||||||||
Db     781 GGGGACAGTGGTTCTCCGATGTTTATTTATGATGCTGAAAAACAAAAATGGTTAATTAAT 840

Qy     853 ggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttccagctggttcgt 912
           ||| |||| |     |||||||| | |  ||| |||   ||||| || ||  |||||||
```


Db 841 GGGATATTACGGGAAGGCAACCCTTTTGAAGGCAAAGAAAATGGGTTTCAATTGGTTCGC 900
Qy 913 aaagattggttctatgatgaaatctttgctggagatacccattcagtattctacgaacca 972
Db 901 AAATCTTATTT---TGATGAAATTTTCGAAAGAGATTTACATACATCACTTTACACCCGA 957
Qy 973 cgtcaaaatgggaaatactcttttaacgacgataataatggcacaggaaaaatcaatgcc 1032
Db 958 GCTGGTAATGGAGTGTACACAATTAGTGGAAATGATAATGGTCAGGGGTCTATAACTCAG 1017
Qy 1033 aaacatgaacacaattctctgcctaatagattaaaaacacgaaccgttcaattgtttaat 1092
Db 1018 AAATCAGGAATACCATCAGAAATTAAAA----------TTACGTTAGCAAAT 1059
Qy 1093 gtttctttatccgagacagcaagagaacctgtttatcatgctgca---ggtggtgtcaac 1149
Db 1060 ATGAGTTTACCTTTGAAAGAGAAGGATAAAGTTCATAATCCTAGATATGACGGACCTAAT 1119
Qy 1150 agttatcgacccagactgaataatggagaaaatatttccttttattgacgaaggaaaaggc 1209
Db 1120 ATTTATTCTCCACGTTTAAACAATGGAGAAACGCTATATTTTATGGATCAAAAACAAGGA 1179
Qy 1210 gaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttccaaggagat 1269
Db 1180 TCATTAATCTTCGCATCTGACATTAACCAAGGGGCGGGTGGTCTTTATTTTGAGGGTAAT 1239
Qy 1270 tttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcatatcagtgaa 1329
Db 1240 TTTACAGTATCTCCAAATTCTAACCAAACTTGGCAAGGAGCTGGCATACATGTAAGTGAA 1299
Qy 1330 gacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggc 1389
Db 1300 AATAGCACCGTTACTTGGAAAGTAAATGGCGTGGAACATGATCGACTTTCTAAAATTGGT 1359
Qy 1390 aaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagcgtgggcgac 1449
Db 1360 AAAGGAACATTGCACGTTCAAGCCAAAGGGGAAAATAAAGGTTCGATCAGCGTAGGCGAT 1419
Qy 1450 ggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcctttagtgaa 1509
Db 1420 GGTAAAGTCATTTTGGAGCAGCAGGCAGACGATCAAGGCAACAAACAAGCCTTTAGTGAA 1479
Qy 1510 atcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcagttcaacccc 1569
Db 1480 ATTGGCTTGGTTAGCGGCAGAGGGACTGTTCAATTAAACGATGATAAACAATTTGATACC 1539
Qy 1570 gacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcattcgctttcg 1629
Db 1540 GATAAATTTTATTTCGGCTTTCGTGGTGGTCGCTTAGATCTTAACGGGCATTCATTAACC 1599
Qy 1630 ttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaa 1689
Db 1600 TTTAAACGTATCCAAAATACGGACGAGGGGGCAATGATTGTGAACCATAATACAACTCAA 1659
Qy 1690 gaatccaccgttaccattacaggcaataaagatattg------ctacaaccggcaataac 1743
Db 1660 GCCGCTAATGTCACTATTACTGGGAACGAAAGCATTGTTCTACCTAATGGAAATAATATT 1719
Qy 1744 aacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagatacg 1803
Db 1720 AATAAACTTGATTACAGAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAAACAGATAAA 1779
Qy 1804 accaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcaccctg 1863
Db 1780 AATAAACACAATGGGCGATTAAACCTTATTTATAAACCAACCACAGAAGATCGTACTTTG 1839
Qy 1864 ctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactgttt 1923
Db 1840 CTACTTTCAGGTGGTACAAATTTAAAAGGCGATATTACCCAAACAAAAGGTAAACTATTT 1899
Qy 1924 ttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaaaaa 1983
Db 1900 TTCAGCGGTAGACCGACACCGCACGCCTACAATCATTTAAATAAACGTTGGTCAGAAATG 1959
Qy 1984 gagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacatttaaa 2043
Db 1960 GAAGGTATACCACAAGGCGAAATTGTGTGGGATCACGATTGGATCAACCGTACATTTAAA 2019
Qy 2044 gcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaagtg 2103
Db 2020 GCTGAAAACTTCCAAATTAAAGGCGGAAGTGCGGTGGTTTCTCGCAATGTTTCTTCAATT 2079
Qy 2104 aaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcatcaa 2163
Db 2080 GAGGGAAATTGGACAGTCAGCAATAATGCAAATGCCACATTTGGTGTTGTGCCAAATCAA 2139
Qy 2164 agccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaaacc 2223
Db 2140 CAAAATACCATTTGCACGCGTTCAGATTGGACAGGATTAACGACTTGTCAAAAAGTGGAT 2199
Qy 2224 attaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtcgat 2283
Db 2200 TTAACCGATACAAAAGTTATTAATTCTATACCAAAAACACAAATCAATGGCTCTATTAAT 2259
Qy 2284 cttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaatcttagt 2343
Db 2260 TTAACTGATAATGCAACGGCGAATGTTAAAGGTTTAGCAAAACTTAATGGCAATGTCACT 2319
Qy 2344 gcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaaccttagc 2403
Db 2320 TTAACAAATCACAGCCAATTTACATTAAGCAACAATGCCACCCAAATAGGCAATATTCG- 2378
Qy 2404 ctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcggct 2463
Db 2379 ---------- 2378
Qy 2464 tcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacgctt 2523
Db 2379 ----------ACTT 2382
Qy 2524 tccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctccctagcc 2583
Db 2383 TCCGACAATTCAACTGCAACGGTGGATAATGCAAACTTGAACGGTAATGTGCATTTAACG 2442
Qy 2584 gataaggcagtattccattttgaaagcagccgctttaccggacaaatcagcggcggcaag 2643
Db 2443 GATTCAGCTCAATTTTCTTTAAAAAACAGCCATTTTTCGCACCAAATTCAGGGAGACAAA 2502
Qy 2644 gatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaattaggc 2703
Db 2503 GGCACAACAGTGACGTTGGAAAATGCGACTTGGACAATGCCTAGCGATACTACATTGCAG 2562
Qy 2704 aatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcggca 2763
Db 2563 AATTTAACGCTAAATAACAGTACGATCACGTTAAATTCAGCTTATT---------- 2608
Qy 2764 ggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgccgt 2823
Db 2609 ----------CAGCTAGCTCAAACAATACGCCACGTCGCCGTTCAT---------- 2644
Qy 2824 tccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacggta 2883
Db 2645 ----TAGAGACGGAAACAACGCCAACATCGGCAGAACATCGTTTCAACACATTGACAGTA 2700
Qy 2884 aacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctaccgc 2943
Db 2701 AATGGTAAATTGAGTGGGCAAGGCACATTCCAATTTACTTCATCTTTATTTGGCTATAAA 2760
Qy 2944 agcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaacaat 3003
Db 2761 AGCGATAAATTAAAATTATCCAATGACGCTGAGGGCGATTACATATTATCTGTTCGCAAC 2820
Qy 3004 accggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaacaaa 3063
Db 2821 ACAGGCAAAGAACCCGAAACCCTTGAGCAATTAACTTTGGTTGAAAGCAAAGATAATCAA 2880

```
Qy   3064 ccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcgtgg 3123
          ||| | || || || || || || || ||  |  |||| || ||||| ||||| || || |
Db   2881 CCGTTATCAGATAAGCTCAAATTTACTTTAGAAAATGACCACGTTGATGCAGGTGCATTA 2940

Qy   3124 cgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaa 3183
          |||||  || |  |    || || ||||| |||||| ||||||| ||  | |||||| ||
Db   2941 CGTTATAAATTAGTGAAGAATGATGGCGAATTCCGCTTGCATAACCCAATAAAAGAGCAG 3000

Qy   3184 gagctttccgacaaactcggcaaggc--------------------agaagccaaaaaa 3222
          ||  |   | |  | | | | ||                     ||||||||||| ||
Db   3001 GAATTGCACAATGATTTAGTAAGAGCAGAGCAAGCAGAACGAACATTAGAAGCCAAACAA 3060

Qy   3223 caggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgcc 3282
          |     |  |  ||  | ||||    ||| |    |  | |  |  | |  ||
Db   3061 GTTGAACCGACTGCTAAAACACAAACAGGTGAGCCAAAAGTGCGGTCAAGAAGAGCAGCG 3120

Qy   3283 gtcgaaaag---acagaaagcgttgccgaaccggcccggcaggcaggcggggaaaatgtc 3339
          | |  |     | || | | |  | || |    || |      |    |||
Db   3121 AGAGCAGCGTTTCCTGATACCCTGCCTGATCAAAGCCTGTTAAACGCATTAGAAGCCAAA 3180

Qy   3340 ggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaagacaccgccttg 3399
          |   | | | |  |||    |||||  |   |          ||| |       |
Db   3181 CAAGCTGAACTGACTGCTGAAACACAAAAAAGTAAGGCAAAAACAAAAAAAGTGCGGTCA 3240

Qy   3400 gcgaaacagcgcgaagcggaaacccggccggctaccaccgccttcccccgcgcccgccgc 3459
          | ||       | ||  ||| ||  | |   | |   ||| |     |     |
Db   3241 AAAAGAGCAGTGTTTTCTGATCCCCTGCTTGATCAAAGCCTGTTCGCATTAGAAGCCGCA 3300

Qy   3460 gcccgccgggatttgccgcaactgcaaccccaaccgcagccccaacc------------- 3506
               |    | | || |||| |  ||    | |  | | |
Db   3301 CTTGAGGTTATTGATGCCCCACAGCAATCGGAAAAAGATCGTCTAGCTCAAGAAGAAGCG 3360

Qy   3507 -------------gcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaa 3552
                        ||    ||| |||||||||||||| | |||||  |  ||    |||
Db   3361 GAAAAACAACGCAAACAAAAAGACTTGATCAGCCGTTATTCAAATAGTGCGTTATCAGAA 3420

Qy   3553 ttttccgccacgctcaacagcgttttcgccgtacaggacgaattagaccgcgtatttgcc 3612
          || || || ||  | || ||  |  |  | || || || |||||||| ||  | ||||
Db   3421 TTATCTGCAACAGTAAATAGTATGCTTTCTGTTCAAGATGAATTAGATCGTCTTTTTGTA 3480

Qy   3613 gaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcg 3672
          ||  |   |     ||||| |||||||    |   | ||| |  | || |||   |||
Db   3481 GATCAAGCACAATCTGCCGTGTGGACAAATATCGCACAGGATAAAAGACGCTATGATTCT 3540

Qy   3673 caagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaac 3732
          | |  ||||| || || |  ||  |||| || | || ||||| ||  |||| |||  |
Db   3541 GATGCGTTCCGTGCTTATCAGCAGAAAACGAACTTACGTCAAATTGGGGTGCAAAAAGCC 3600

Qy   3733 ctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaaaacaccttcgac 3792
          | |  |  || ||  | ||     | || ||||| | |||  | || || ||||| ||
Db   3601 TTAGCTAATGGACGAATTGGGGCAGTTTTCTCGCATAGCCGTTCAGATAATACCTTTGAT 3660

Qy   3793 gacggcatcggcaactcggcacggcttgcccacggcgccgttttcgggcaatacggcatc 3852
          ||    |   ||   ||    |  |        | | ||| |  |||||
Db   3661 GAACAGGTTAAAAATCACGCGACATTAACGATGATGTCGGGTTTTGCCCAATATCAATGG 3720

Qy   3853 gacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggcagcctttcagac 3912
          | |   ||  |  | ||  | | || |||  ||||  | ||     | |   |  | ||
Db   3721 GGCGATTTACAATTTGGTGTAAACGTGGGAACGGGAATCAGTGCGAGTAAAATGGCTGAA 3780

Qy   3913 ggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcaggcacgataccgc 3972
          |       |  ||||| |  ||    | |  |||| ||| |  | ||| | ||  | ||
Db   3781 GAACAAAGCCGAAAAATTCATCGAAAAGCGATAAATTATGGCGTGAATGCAAGTTATCAG 3840

Qy   3973 gccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtccaa 4032
          | |||| ||   ||| |||||  | ||  |  | || |  |  |||||||  |  ||
Db   3841 TTCCGTTTAGGGCAATTGGGCATTCAGCCTTATTTTGGAGTTAATCGCTATTTTATTGAA 3900

Qy   4033 aaagcggattaccgctacgaaaacgtcaatatcgccacccccggccttgcattcaaccgc 4092
          |  |||| |  |  ||  | || |    |    || ||  |||||||||| || |||
Db   3901 CGTGAAAATTATCAATCTGAGGAAGTGAGAGTGAAAACGCCTAGCCTTGCATTTAATCGC 3960

Qy   4093 taccgcgcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatcacg 4152
          ||    || ||||||  |  |||||| |||| |  ||| |  |  | ||   | | | |
Db   3961 TATAATGCTGGCATTCGAGTTGATTATACATTTACTCCGACAGATAATATCAGCGTTAAG 4020

Qy   4153 ccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccgaacacgcgtcaat 4212
          ||||||||  | |   |||  |||| |  |  || || | |||  |||||| |||
Db   4021 CCTTATTTCTTCGTCAATTATGTTGATGTTTCAAACGCTAACGTACAAACCACGGTAAAT 4080

Qy   4213 accgccgtattggctcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgcc 4272
          | | || |||   ||   || ||         |       |||  |||  |||| ||
Db   4081 CTCACGGTGTTGCAACAACCATTTGGACGTTATTGGCAAAAAGAAGTGGGATTAAAGGCA 4140

Qy   4273 gaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggcccgcaactggaa 4332
          |||||  |  ||||   | |||          |    |  ||||  | ||||| |
Db   4141 GAAATTTTACATTTCCAAATTTCCGCTTTTATCTCAAAATCTCAAGGTTCACAACTCGGC 4200

Qy   4333 gcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
          || || |  | |||| | ||||| ||||| || ||||||
Db   4201 AAACAGCAAAATGTGGGCGTGAAATTGGGCTATCGTTGGTAA 4242


RESULT  14
AR157320
LOCUS       AR157320                4319 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6245337.
ACCESSION   AR157320
VERSION     AR157320.1  GI:16218252
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4319)
  AUTHORS   St. Geme,J.W. III and Falkow,S.
  TITLE     Haemophilus adherence and penetration proteins
  JOURNAL   Patent: US 6245337-A 1 12-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .4319
                     /organism="unknown"
BASE COUNT     1497 a    776 c    891 g   1155 t
ORIGIN


  Query Match           28.6%;  Score 1249.4;  DB 6;  Length 4319;
  Best Local Similarity   58.8%;  Pred. No. 8e-271;
  Matches 2565;  Conservative    0;  Mismatches 1536;  Indels  264;  Gaps   13;

Qy      73 cctgcttacttagccatatgcctgtcgttcggcattcttccccaagcctgggcgggacac 132
           | |  ||  ||| ||   ||| | || || || ||  |  | ||||| ||||| || |||
Db      81 CTTAATTTTTTAACCGCTTGCATTTCATTAGGGATAGTATCGCAAGCGTGGGCTGGTCAC 140

Qy     133 acttatttcggcatcaactaccaatactatcgcgactttgccgaaaataaaggcaagttt 192
           |||||||| || ||  | |||||||| ||||| || |||||||| |||||||| |||||
Db     141 ACTTATTTTGGGATTGATTACCAATATTATCGTGATTTTGCCGAGAATAAAGGGAAGTTC 200

Qy     193 gcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatca 252
           |||| |||||  || ||||| ||||||| |||||| ||||| | || || |||| ||||
Db     201 ACAGTTGGGGCTCAAAATATTAAGGTTTATAACAAACAAGGGCAATTAGTTGGCACATCA 260

Qy     253 atgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattg 312
           ||||||||||||||||||||||||||||||| |||| |||||||||||||| || |||
Db     261 ATGACAAAAGCCCCGATGATTGATTTTTCTGTAGTGTCACGTAACGGCGTGGCAGCCTTG 320

Qy     313 gtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgatttt 372
           || |  |||||||||||||||||||||||||||||  || ||||  | ||||||||
Db     321 GTTGAAAATCAATATATTGTGAGCGTGGCACATAACGTAGGATATACAGATGTTGATTTT 380

Qy     373 ggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaattgtgaaacggaat 432
           |||||  || ||||  || |||||||||||||||||||||||||| ||||| ||||| |||
Db     381 GGTGCAGAGGGAAACAACCCCGATCAACATCGTTTTACTTATAAGATTGTAAAACGAAAT 440

Qy     433 aattataaagcagggactaaaggccatccttatggcggcgattatcatatgccgcgtttg 492
           || || |||  ||    |||    ||||||||||  | |||||| |||  || || ||
Db     441 AACTACAAAAAAG---ATAATTTACATCCTTATGAGGACGATTACCATAATCCACGATTA 497
```

```
Qy    493 cataaatttgtcacagatgcagaacctgttgaaatgaccagttatatggatgggcggaaa 552
          ||||||||| || ||||| || | || |||| |||||     ||||| ||||  | |
Db    498 CATAAATTCGTTACAGAAGCGGCTCCAATTGATATGACTTCGAATATGAATGGCAGTACT 557

Qy    553 tatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggcaggcaatattgg 612
          |||   |||  ||  || || || || ||||||||||| ||  | ||  |||| | ||||
Db    558 TATTCAGATAGAACAAAATATCCAGAACGTGTTCGTATCGGCTCTGGACGGCAGTTTTGG 617

Qy    613 cgatctgatgaagatgagcccaataaccgcgaaagttcatatcatattgcaagtgcgtat 672
          |||  |||| ||||  |   |                    | ||  |||| |||| |||
Db    618 CGAAATGATCAAGACAAAGGC------------------GACCAAGTTGCCGGTGCATAT 659

Qy    673 tcttggctcgttggtggcaatacctttgcacaaaatggatcaggtggtggcacagtcaac 732
            ||  ||    | |||||||||       ||   |||| |||||  |||       |
Db    660 CATTATCTGACAGCTGGCAATACACACAATCAGCGTGGAGCAGGTAATGGATATTCGTAT 719

Qy    733 ttaggtagtgaaaaaattaaacatagcccatatggtttttttaccaacaggaggctcattt 792
          || ||  | ||      ||||    |  |||||||   ||||| |  | |||||||
Db    720 TTGGGAGGCGATGTTCGTAAAGCGGGAGAATATGGTCCATTACCGATTGCAGGCTCAAAG 779

Qy    793 ggcgacagtggctcaccaatgtttatctatgatgcccaaaagcaaaagtggttaattaat 852
          || |||||||| || || |||||||| ||||||||  |||| |||||  ||||||||||||
Db    780 GGGGACAGTGGTTCTCCGATGTTTATTTATGATGCTGAAAAACAAAAATGGTTAATTAAT 839

Qy    853 ggggtattgcaaacgggcaacccctatataggaaaaagcaatggcttccagctggttcgt 912
          ||| |||| |      |||||||| | |  ||| |||   ||||| || ||  |||||||
Db    840 GGGATATTACGGGAAGGCAACCCTTTTGAAGGCAAAGAAAATGGGTTTCAATTGGTTCGC 899

Qy    913 aaagattggttctatgatgaaatctttgctggagatacccattcagtattctacgaacca 972
          |||  ||  ||   ||||||||| || |   |||||   ||| ||  | | |||   | |
Db    900 AAATCTTATTT---TGATGAAATTTTCGAAAGAGATTTACATACATCACTTTACACCCGA 956

Qy    973 cgtcaaaatgggaaatactcttttaacgacgataataatggcacaggaaaaatcaatgcc 1032
            |  |||||   ||| |  |||  |   || |||||||    ||    || | |
Db    957 GCTGGTAATGGAGTGTACACAATTAGTGGAAATGATAATGGTCAGGGGTCTATAACTCAG 1016

Qy   1033 aaacatgaacacaattctctgcctaatagattaaaaacacgaaccgttcaattgtttaat 1092
          |||   | |      ||      ||| |                    ||   ||  |||
Db   1017 AAATCAGGAATACCATCAGAAATTAAAA-----------------TTACGTTAGCAAAT 1058

Qy   1093 gtttctttatccgagacagcaagagaacctgtttatcatgctgca---ggtggtgtcaac 1149
            |  |||| |   || ||  |  ||    ||| || || ||  |   |  ||    ||
Db   1059 ATGAGTTTACCTTTGAAAGAGAAGGATAAAGTTCATAATCCTAGATATGACGGACCTAAT 1118

Qy   1150 agttatcgacccagactgaataatggagaaaatatttccttttattgacgaaggaaaaggc 1209
          | ||||   ||  |  | || |||||||||   | |  ||||| ||  ||  | ||||
Db   1119 ATTTATTCTCCACGTTTAAACAATGGAGAAACGCTATATTTTATGGATCAAAAACAAGGA 1178

Qy   1210 gaattgatacttaccagcaacatcaatcaaggtgctggaggattatatttccaaggagat 1269
            ||| ||  |  |     |||| || ||||| || || ||  |.|||||  | ||  ||
Db   1179 TCATTAATCTTCGCATCTGACATTAACCAAGGGGCGGGTGGTCTTTATTTTGAGGGTAAT 1238

Qy   1270 tttacggtctcgcctgaaaataacgaaacttggcaaggcgcgggcgttcatatcagtgaa 1329
          ||||| || || ||  |   |||| |||||||||||| || ||| | ||| | ||||||
Db   1239 TTTACAGTATCTCCAAATTCTAACCAAACTTGGCAAGGAGCTGGCATACATGTAAGTGAA 1298

Qy   1330 gacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggc 1389
            | || |||||||||||||||||| ||||||| | | || || || || ||||| ||
Db   1299 AATAGCACCGTTACTTGGAAAGTAAATGGCGTGGAACATGATCGACTTTCTAAAATTGGT 1358

Qy   1390 aaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagcgtgggcgac 1449
          |||||  ||  |||||||||||||||||||||||  |||| |||||||||| |||||
Db   1359 AAAGGAACATTGCACGTTCAAGCCAAAGGGGAAAATAAAGGTTCGATCAGCGTAGGCGAT 1418

Qy   1450 ggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcctttagtgaa 1509
          |||| |||||||||||| |||||||||||||| ||||||| |||||||||||||||||||
Db   1419 GGTAAAGTCATTTTGGAGCAGCAGGCAGACGATCAAGGCAACAAACAAGCCTTTAGTGAA 1478

Qy   1510 atcggcttggtcagcggcaggggtacggtgcaactgaatgccgataatcagttcaacccc 1569
          || |||||||| |||||||| || || || ||| | || |  |||| || ||  |  ||
Db   1479 ATTGGCTTGGTTAGCGGCAGAGGGACTGTTCAATTAAACGATGATAAACAATTTGATACC 1538

Qy   1570 gacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcattcgctttcg 1629
```

```
                || ||| | |||||||||||||| || || || || ||| | ||||||||||  |  |
Db   1539 GATAAATTTTATTTCGGCTTTCGTGGTGGTCGCTTAGATCTTAACGGGCATTCATTAACC 1598

Qy   1630 ttccaccgtattcaaaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaa 1689
          ||  | ||||| |||||||| || || ||||| |||||||| ||||| |||  |    ||
Db   1599 TTTAAACGTATCCAAAATACGGACGAGGGGGCAATGATTGTGAACCATAATACAACTCAA 1658

Qy   1690 gaatccaccgttaccattacaggcaataaagatattg------ctacaaccggcaataac 1743
          |   | |  || || |||||  || ||  ||   ||||      |||        ||||
Db   1659 GCCGCTAATGTCACTATTACTGGGAACGAAAGCATTGTTCTACCTAATGGAAATAATATT 1718

Qy   1744 aacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgagaaagatacg 1803
          || |  | |||  || ||||||||||||||||||||||||||||||||||| | |||||
Db   1719 AATAAACTTGATTACAGAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAAACAGATAAA 1778

Qy   1804 accaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcaccctg 1863
          |  |||   || ||||| | |||||| |||| | ||  || ||||||| || ||  ||
Db   1779 AATAAACACAATGGGCGATTAAACCTTATTTATAAACCAACCACAGAAGATCGTACTTTG 1838

Qy   1864 ctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactgttt 1923
          || ||||| || || |||||||||||| ||| | || || |||||||| || ||||| |||
Db   1839 CTACTTTCAGGTGGTACAAATTTAAAAGGCGATATTACCCAAACAAAAGGTAAACTATTT 1898

Qy   1924 ttcagcggcagaccaacaccgcacgcctacaatcatttaaacgaccattggtcgcaaaaa 1983
          |||||||| ||||| |||||||||||||||||||||||||||  | | ||||||  |||
Db   1899 TTCAGCGGTAGACCGACACCGCACGCCTACAATCATTTAAATAAACGTTGGTCAGAAATG 1958

Qy   1984 gagggcattcctcgcggggaaatcgtgtgggacaacgactggatcaaccgcacatttaaa 2043
          || || || || |  || ||||| |||||||| |||| ||||||||||| |||||||||
Db   1959 GAAGGTATACCACAAGGCGAAATTGTGTGGGATCACGATTGGATCAACCGTACATTTAAA 2018

Qy   2044 gcggaaaacttccaaattaaaggcggacaggcggtggtttcccgcaatgttgccaaagtg 2103
          || |||||||||||||||||||||||   |||||||||| |||||||||  |  | |
Db   2019 GCTGAAAACTTCCAAATTAAAGGCGGAAGTGCGGTGGTTTCTCGCAATGTTTCTTCAATT 2078

Qy   2104 aaaggcgattggcatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcatcaa 2163
            | ||  |||||    | |||||| | ||  | ||    |||||||| |  ||  |||||
Db   2079 GAGGGAAATTGGACAGTCAGCAATAATGCAAATGCCACATTTGGTGTTGTGCCAAATCAA 2138

Qy   2164 agccacacaatctgtacacgttcggactggacgggtctgacaaattgtgtcgaaaaaacc 2223
              | || || ||  || ||||| || ||||| ||  | || | ||||    ||
Db   2139 CAAAATACCATTTGCACGCGTTCAGATTGGACAGGATTAACGACTTGTCAAAAAGTGGAT 2198

Qy   2224 attaccgacgataaagtgattgcttcattgactaagaccgacatcagcggcaatgtcgat 2283
          | |||||    ||||| |||   |||  |  || ||  | |||| |||  | |  ||
Db   2199 TTAACCGATACAAAAGTTATTAATTCTATACCAAAAACACAAATCAATGGCTCTATTAAT 2258

Qy   2284 cttgccgatcacgctcatttaaatctcacagggcttgccacactcaacggcaatcttagt 2343
          |  | ||| | ||       ||| | | |||  | || | ||| || |||||| | | |
Db   2259 TTAACTGATAATGCAACGGCGAATGTTAAAGGTTTAGCAAAACTTAATGGCAATGTCACT 2318

Qy   2344 gcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggcaaccttagc 2403
            ||      | |  |  | |||| | ||| |||| |||||||||| |||||  || |
Db   2319 TTAACAAATCACAGCCAATTTACATTAAGCAACAATGCCACCCAAATAGGCAATATTCG- 2377

Qy   2404 ctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaacacatcggct 2463

Db   2378 ------------------------------------------------------------ 2377

Qy   2464 tcgggcaatgcttcatttaatctaagcgaccacgccgtacaaaacggcagtctgacgctt 2523
                                                                   |||
Db   2378 ----------------------------------------------------------ACTT 2381

Qy   2524 tccggcaacgctaaggcaaacgtaagccattccgcactcaacggtaatgtctccctagcc 2583
          |||| ||| | | |||| ||    ||  |     | ||||||||||     || |
Db   2382 TCCGACAATTCAACTGCAACGGTGGATAATGCAAACTTGAACGGTAATGTGCATTTAACG 2441

Qy   2584 gataaggcagtattccatttgaaagcagccgctttaccggacaaatcagcggcggcaag 2643
          |||   ||  |||  ||| ||| ||||| ||| |    |||||    || | |||
Db   2442 GATTCAGCTCAATTTTCTTTAAAAAACAGCCATTTTTCGCACCAAATTCAGGGAGACAAA 2501

Qy   2644 gatacggcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaattaggc 2703
          |  ||  || |    ||  || |       |||||  ||||    |  ||   |||
```

```
Db  2502 GGCACAACAGTGACGTTGGAAAATGCGACTTGGACAATGCCTAGCGATACTACATTGCAG 2561
Qy  2704 aatttaaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgcggca 2763
Db  2562 AATTTAACGCTAAATAACAGTACGATCACGTTAAATTCAGCTTATT-------------- 2607
Qy  2764 ggggcgcaaaccggcagtgcgacagatgcgccgcgccgccgttcgcgccgttcgcgccgt 2823
Db  2608 ----------CAGCTAGCTCAAACAATACGCCACGTCGCCGTTCA-------------- 2642
Qy  2824 tccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacggta 2883
Db  2643 ---TTAGAGACGGAAACAACGCCAACATCGGCAGAACATCGTTTCAACACATTGACAGTA 2699
Qy  2884 aacggcaaattgaacggtcagggaacattccgctttatgtcggaactcttcggctaccgc 2943
Db  2700 AATGGTAAATTGAGTGGGCAAGGCACATTCCAATTTACTTCATCTTTATTTGGCTATAAA 2759
Qy  2944 agcgacaaattgaagctggcggaaagttccgaaggcacttacaccttggcggtcaacaat 3003
Db  2760 AGCGATAAATTAAAATTATCCAATGACGCTGAGGGCGATTACATATTATCTGTTCGCAAC 2819
Qy  3004 accggcaacgaacctgcaagcctcgaacaattgacggtagtggaaggaaaagacaacaaa 3063
Db  2820 ACAGGCAAAGAACCCGAAACCCTTGAGCAATTAACTTTGGTTGAAAGCAAAGATAATCAA 2879
Qy  3064 ccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcgtgg 3123
Db  2880 CCGTTATCAGATAAGCTCAAATTTACTTTAGAAAATGACCACGTTGATGCAGGTGCATTA 2939
Qy  3124 cgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaa 3183
Db  2940 CGTTATAAATTAGTGAAGAATGATGGCGAATTCCGCTTGCATAACCCAATAAAAGAGCAG 2999
Qy  3184 gagctttccgacaaactcggcaaggc----------------------agaaggccaaaaaa 3222
Db  3000 GAATTGCACAATGATTTAGTAAGAGCAGAGCAAGCAGAACGAACATTAGAAGCCAAACAA 3059
Qy  3223 caggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgcc 3282
Db  3060 GTTGAACCGACTGCTAAAACACAAACAGGTGAGCCAAAAGTGCGGTCAAGAAGAGCAGCG 3119
Qy  3283 gtcgaaaag---acagaaagcgttgccgaaccggcccggcaggcaggcggggaaaatgtc 3339
Db  3120 AGAGCAGCGTTTCCTGATACCCTGCCTGATCAAAGCCTGTTAAACGCATTAGAAGCCAAA 3179
Qy  3340 ggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaagacaccgccttg 3399
Db  3180 CAAGCTGAACTGACTGCTGAAACACAAAAAAGTAAGGCAAAAACAAAAAAAGTGCGGTCA 3239
Qy  3400 gcgaaacagcgcgaagcggaaacccggccggctaccaccgccttcccccgcgcccgccgc 3459
Db  3240 AAAAGAGCAGTGTTTTCTGATCCCCTGCTTGATCAAAGCCTGTTCGCATTAGAAGCCGCA 3299
Qy  3460 gcccgccgggatttgccgcaactgcaaccccaaccgcagccccaacc-------------- 3506
Db  3300 CTTGAGGTTATTGATGCCCCACAGCAATCGGAAAAAGATCGTCTAGCTCAAGAAGAAGCG 3359
Qy  3507 --------------gcagcgcgacctgatcagccgttatgccaatagcggtttgagtgaa 3552
Db  3360 GAAAAACAACGCAAACAAAAAGACTTGATCAGCCGTTATTCAAATAGTGCGTTATCAGAA 3419
Qy  3553 ttttccgccacgctcaacagcgttttcgccgtacaggacgaattagaccgcgtatttgcc 3612
Db  3420 TTATCTGCAACAGTAAATAGTATGCTTTCTGTTCAAGATGAATTAGATCGTCTTTTTGTA 3479
Qy  3613 gaagaccgccgcaacgccgtttggacaagcggcatccgggacaccaaacactaccgttcg 3672
Db  3480 GATCAAGCACAATCTGCCGTGTGGACAAATATCGCACAGGATAAAAGACGCTATGATTCT 3539
Qy  3673 caagatttccgcgcctaccgccaac---aaaccgacctgcgccaaatcggtatgcagaaa 3729
Db  3540 GATGCGTTCCGTGCTTATCAGCAGCAGAAAACGAACTTACGTCAAATTGGGGTGCAAAAA 3599
```

```
Qy  3730 aacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaaaacaccttc 3789
Db  3600 GCCTTAGCTAATGGACGAATTGGGGCAGTTTTCTCGCATAGCCGTTCAGATAATACCTTT 3659
Qy  3790 gacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcgggcaatacggc 3849
Db  3660 GATGAACAGGTTAAAAATCACGCGACATTAACGATGATGTCGGGTTTTGCCCAATATCAA 3719
Qy  3850 atcgacaggttctacatcggcatcagcgcgggcgcgggttttagcagcggcagcctttca 3909
Db  3720 TGGGGCGATTTACAATTTGGTGTAAACGTGGGAACGGGAATCAGTGCGAGTAAAATGGCT 3779
Qy  3910 gacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattcaggcacgatac 3969
Db  3780 GAAGAACAAAGCCGAAAAATTCATCGAAAAGCGATAAATTATGGCGTGAATGCAAGTTAT 3839
Qy  3970 cgcgccggtttcggcggattcggcatcgaaccgcacatcggcgcaacgcgctatttcgtc 4029
Db  3840 CAGTTCCGTTTAGGGCAATTGGGCATTCAGCCTTATTTTGGAGTTAATCGCTATTTTATT 3899
Qy  4030 caaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggccttgcattcaac 4089
Db  3900 GAACGTGAAAATTATCAATCTGAGGAAGTGAGAGTGAAAACGCCTAGCCTTGCATTTAAT 3959
Qy  4090 cgctaccgcgcgggcattaaggcagattattcattcaaaccggcgcaacacatttccatc 4149
Db  3960 CGCTATAATGCTGGCATTCGAGTTGATTATACATTTACTCCGACAGATAATATCAGCGTT 4019
Qy  4150 acgccttatttgagcctgtcctataccgatgccgcttcgggcaaagtccgaacacgcgtc 4209
Db  4020 AAGCCTTATTTCTTCGTCAATTATGTTGATGTTTCAAACGCTAACGTACAAACCACGGTA 4079
Qy  4210 aataccgccgtattggctcaggatttcggcaaaacccgcagtgcggaatggggcgtaaac 4269
Db  4080 AATCTCACGGTGTTGCAACAACCATTTGGACGTTATTGGCAAAAAGAAGTGGGATTAAAG 4139
Qy  4270 gccgaaatcaaaggtttcacgctgtccctccacgctgccgccgccaaaggcccgcaactg 4329
Db  4140 GCAGAAATTTTACATTTCCAAATTTCCGCTTTTATCTCAAAATCTCAAGGTTCACAACTC 4199
Qy  4330 gaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4374
Db  4200 GGCAAACAGCAAAATGTGGGCGTGAAATTGGGCTATCGTTGGTAA 4244
```

```
RESULT  15
NLIGA3327
LOCUS       NLIGA3327               578 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3327.
ACCESSION   AJ001740
VERSION     AJ001740.1  GI:3152399
KEYWORDS    IgA protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 578)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Neisseria lactamica"
                     /isolate="NL3327"
                     /db_xref="taxon:486"
```

```
     CDS             <1. .>576
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA04965.1"
                     /db_xref="GI:3152400"
                     /db_xref="SPTREMBL:O69746"
                     /translation="GKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSYVSRNGVAALV
                     GDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRL
                     HKFVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGAWLI
                     GGNTHMQGWGNNGVVSLSGDVRHANDYGPMPI"
     repeat_region   88. .95
                     /rpt_type=INVERTED
     repeat_region   139. .148
BASE COUNT      163 a    125 c    156 g    134 t
ORIGIN


  Query Match             7.3%;  Score 318.8;  DB 1;  Length 578;
  Best Local Similarity   83.3%;  Pred. No. 4.1e-61;
  Matches  375;  Conservative    0;  Mismatches   72;  Indels    3;  Gaps    1;

Qy    184 ggcaagtttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtc 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 60

Qy    244 ggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtg 303
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     61 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTG 120

Qy    304 gcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaac 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAAT 180

Qy    364 gttgattttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaattgtg 423
          |||||||||||||| || ||||| |||||||||||| || || ||| |||| ||||||||
Db    181 GTTGATTTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG 240

Qy    424 aaacggaataattataaagcagggactaaaggccatccttatggcggcgattatcatatg 483
          ||| | ||||| || ||| | |  | |  |   || |||||   ||||||||| ||||||
Db    241 AAAAGAAATAACTACAAACCTGACAATTCA---CACCCTTACAACGGCGATTACCATATG 297

Qy    484 ccgcgtttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatggat 543
          |||||  ||||||| || || || ||||||||||||||| ||||||||  || | |||
Db    298 CCGCGCCTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGG 357

Qy    544 gggcggaaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggcagg 603
          |||   |  |||  |||| || | || || || || ||||| || || ||  ||||
Db    358 GGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACAC 417

Qy    604 caatattggcgatctgatgaagatgagccc 633
          || |||||||| | |||||| ||  | | |
Db    418 CACTATTGGCGTTATGATGATGACAAACAC 447


RESULT  16
NLIGA3354
LOCUS       NLIGA3354                584 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3354.
ACCESSION   AJ001739
VERSION     AJ001739.1  GI:3152397
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 584)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Neisseria lactamica"
                     /isolate="NL3354"
                     /db_xref="taxon:486"
     CDS             <1. .>582
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04964.1"
                     /db_xref="GI:3152398"
                     /db_xref="SPTREMBL:O69747"
                     /translation="FAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGD
                     QYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLHK
                     FVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDDKHGDLSYSGAWLIG
                     GNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGA"
     repeat_region   82. .88
                     /rpt_type=INVERTED
     repeat_region   134. .142
BASE COUNT      163 a    127 c    159 g    135 t
ORIGIN


  Query Match             7.2%;  Score 313.6;  DB 1;  Length 584;
  Best Local Similarity   83.6%;  Pred. No. 6e-60;
  Matches  368;  Conservative    0;  Mismatches   69;  Indels    3;  Gaps    1;

Qy    190 tttgcagtcggggcgaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaa 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAA 60

Qy    250 tcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggca 309
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     61 TCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCA 120

Qy    310 ttggtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgat 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    121 TTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGAT 180

Qy    370 tttggtgcggaaggaagaaatcccgatcaacatcgttttacttataaaattgtgaaacgg 429
          |||||||| || ||||| |||||||||| || || ||| ||||  |||||||||| |
Db    181 TTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGA 240

Qy    430 aataattataaagcagggactaaaggccatccttatggcggcgattatcatatgccgcgt 489
          ||||| || ||| | |  | | |   || |||||  |||||||| ||||||||||
Db    241 AATAACTACAAACCTGACAATTCA---CACCCTTACAACGGCGATTACCATATGCCGCGC 297

Qy    490 ttgcataaatttgtcacagatgcagaacctgttgaaatgaccagttatatggatgggcgg 549
          |||||||  || || || |||||||||||||| ||||||||  || | |||   |||
Db    298 CTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGGAAT 357

Qy    550 aaatatatcgatcaaaataattaccctgaccgtgttcgtattggggcaggcaggcaatat 609
          |  |||  |||| || | || || || || ||||| || || ||  ||||    || |||
Db    358 ACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTAT 417

Qy    610 tggcgatctgatgaagatga 629
          ||||| | |||||| || ||
Db    418 TGGCGTTATGATGATGACGA 437


RESULT  17
NLIGA823
LOCUS       NLIGA823                 572 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL823.
ACCESSION   AJ001737
VERSION     AJ001737.1  GI:3152389
KEYWORDS    IgA protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
```

```
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 572)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Neisseria lactamica"
                     /isolate="NL823"
                     /db_xref="taxon:486"
     CDS             <1. .304
                     /codon_start=2
                     /transl_table=11
                     /product="IgA protease homologue"
                     /protein_id="CAA04962.1"
                     /db_xref="GI:3152390"
                     /db_xref="SPTREMBL:O69748"
                     /translation="KDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGDQYIVS
                     VAHNGGYNNVDFGAEGGNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLPGISS"
     repeat_region   44. .51
                     /rpt_type=INVERTED
     repeat_region   68. .75
BASE COUNT      158 a    124 c    157 g    133 t
ORIGIN
```

```
  Query Match             6.4%;  Score 280.6;  DB 1;  Length 572;
  Best Local Similarity   81.1%;  Pred. No. 1.7e-52;
  Matches  353;  Conservative    0;  Mismatches   74;  Indels    8;  Gaps    2;

Qy    204 gaaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatcaatgacaaaagc 263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 60

Qy    264 cccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattggtgggcgatca 323
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCA 120

Qy    324 atatattgtgagcgtggcacataacggcggctataacaacgttgattttggtgcggaagg 383
          |||||||||||||||||| |||||||||||||||||| |||||||||||||||| ||
Db    121 ATATATTGTGAGCGTGGCGCATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGG 180

Qy    384 aagaaatcccgatcaacatcgttttacttataaaattgtgaaacggaataattataaagc 443
          | | ||||||||||| || || ||| ||||  ||||||| ||| | ||||| || ||| |
Db    181 AGGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTAAAAAGAAATAACTACAAACC 240

Qy    444 agggactaaaggccatccttatggcggcgattatcatatgccgcg-----tttgcataaa 498
          |  | | |  |   || |||||   |||||||| ||||||||||      | ||||||
Db    241 TGACAATTCA---CACCCTTACAACGGCGATTACCATATGCCGCGCCTGCCTGGCATAAG 297

Qy    499 tttgtcacagatgcagaacctgttgaaatgaccagttatatggatgggcggaaatatatc 558
          || || || |||||||||||||| ||||||||  || | |||   |||  |  |||  |
Db    298 TTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGGAATACCTATTCC 357

Qy    559 gatcaaaataattaccctgaccgtgttcgtattggggcaggcaggcaatattggcgatct 618
          ||| || | || || || || ||||| || || ||  ||||    || ||||||| | |
Db    358 GATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTATTGGCGTTAT 417

Qy    619 gatgaagatgagccc 633
          ||||| ||  | | |
Db    418 GATGATGACAAACAC 432
```

```
RESULT  18
NLIGA3293
```

```
LOCUS       NLIGA3293               572 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3293.
ACCESSION   AJ001738
VERSION     AJ001738.1  GI:3152385
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 572)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Neisseria lactamica"
                     /isolate="NL3293"
                     /db_xref="taxon:486"
     CDS             <1. .>570
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04963.1"
                     /db_xref="GI:3152386"
                     /db_xref="SPTREMBL:O69745"
                     /translation="KDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGDQYIVS
                     VAHNVGYRDVDFGAEGSNPDQHRFSYKIAKRNNYKNDETHPYEKDYHNPRLHKFVTEA
                     APIDMTSDMDGNKYTDRTKYPERVRIGSGWQFWRNDQDKGDQVAGAYHYLTAGNTHNQ
                     GGAGGGWSSLSGDVRHAGNYGPIPIAAGSS"
     repeat_region   67. .74
                     /rpt_type=INVERTED
     repeat_region   118. .127
BASE COUNT      169 a    119 c    154 g    130 t
ORIGIN
```

```
  Query Match             6.4%;  Score 280.2;  DB 1;  Length 572;
  Best Local Similarity   77.7%;  Pred. No. 2.1e-52;
  Matches  352;  Conservative    0;  Mismatches   98;  Indels    3;  Gaps    1;

Qy    205 aaagatattgaggtttacaacaaaaaaggggagttggtcggcaaatcaatgacaaaagcc 264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGAAAATCAATGACAAAAGCC 60

Qy    265 ccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattggtgggcgatcaa 324
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     61 CCGATGATTGATTTTTCTGTGGTATCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAA 120

Qy    325 tatattgtgagcgtggcacataacggcggctataacaacgttgattttggtgcggaagga 384
          |||||||||||||||||| ||||| |  || |||  | | ||||||||||||||||||| |||
Db    121 TATATTGTGAGCGTGGCGCATAATGTAGGTTATCGCGATGTTGATTTTGGTGCGGAGGGA 180

Qy    385 agaaatcccgatcaacatcgttttacttataaaattgtgaaacggaataattataaagca 444
          || |||||||||| || |||||| ||||||| |||| ||||||||||||||||||||||
Db    181 AGCAATCCCGATCAGCACCGTTTTTCTTATAAGATTGCAAAACGGAATAATTATAAAAAC 240

Qy    445 gggactaaaggccatccttatggcggcgattatcatatgccgcgtttgcataaatttgtc 504
          |    | ||   ||||||||||      ||||| || |  || || || ||||||||||
Db    241 G---ATGAAACGCATCCTTATGAGAAAGATTACCACAACCCACGCTTACATAAATTTGTT 297

Qy    505 acagatgcagaacctgttgaaatgaccagttatatggatgggcggaaatatatcgatcaa 564
          || || || |  ||  |||| |||||   | ||||||| ||    |||||||  |||
Db    298 ACGGAAGCCGCCCCGATTGATATGACTTCTGATATGGACGGCAACAAATATACAGATAGG 357

Qy    565 aataattaccctgaccgtgttcgtattggggcaggcaggcaatattggcgatctgatgaa 624
          |  || || || || || || ||||| ||  | ||  |||| | |||||||   ||| ||
```

```
Db     358 ACGAAATATCCCGAACGCGTCCGTATCGGCTCCGGGTGGCAGTTTTGGCGAAACGATCAA 417

Qy     625 gatgagcccaataaccgcgaaagttcatatcat 657
           ||  |   | || |    |   |  ||||||||
Db     418 GACAAAGGCGATCAAGTTGCCGGCGCATATCAT 450


RESULT  19
AF012207
LOCUS       AF012207                2992 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z3524 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012207
VERSION     AF012207.1  GI:2529497
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2992
                     /organism="Neisseria meningitidis"
                     /strain="Z3524"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup III; isolated in Chad, 1988"
     mat_peptide     1. .>2992
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2992
                     /gene="iga"
                     /allele="iga3"
     CDS             <1. .>2992
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45790.2"
                     /db_xref="GI:12597967"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     GANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGANNDITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
                     GIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
                     RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGQTENNDWVFMGYKQEEAQKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAR
                     KDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVSNITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFGATQINGNVNLSQNAALVLGKAALW
                     GQIQGQGNSSVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKINH
                     LSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQDR
                     SHLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQKA
                     TQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      986 a    685 c    640 g    681 t
ORIGIN
```

```
  Query Match             4.6%;  Score 202.6;  DB 1;  Length 2992;
  Best Local Similarity   58.3%;  Pred. No. 8.5e-35;
  Matches  374;  Conservative    0;  Mismatches  264;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          ||||  || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db   1035 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1094

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| |||| || |||     || |||||| ||    ||||||| ||| |||||
Db   1095 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGG 1154

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |    || | ||  |  ||   |||| |||| ||      |    |||| |
Db   1155 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1214

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1215 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1274

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || ||||||||||| || ||| || |||| || || || |   |    |||
Db   1275 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1334

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||  | |||||  ||  | | |||
Db   1335 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1394

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| ||  | |||||||| || || ||||| |||||||||   ||
Db   1395 AATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1454

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1455 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1514

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   ||||   |  |||||  | ||  | || |||||   |   ||| |  |  |
Db   1515 CACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1574

Qy   1735 ggcaataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
           |||        |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1575 CAAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTAT 1634

Qy   1795 aaagatacgaccaaaacgaacgggcggctcaaccttgttta 1835
           |       |  ||| | |    ||   |||   | ||||
Db   1635 TACAGCTATAGAAAACCCATTCCGCAAGGCAAAGATCTTTA 1675


RESULT  20
AF012206
LOCUS       AF012206                2992 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4024 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012206
VERSION     AF012206.1  GI:2529495
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
```

```
           Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2992
                     /organism="Neisseria meningitidis"
                     /strain="Z4024"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup VI; isolated in East-Germany,
                     1985"
     mat_peptide     1. .>2992
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2992
                     /gene="iga"
                     /allele="iga6"
     CDS             <1. .>2992
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45789.2"
                     /db_xref="GI:12597966"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     DANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQV
                     GIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTS
                     HASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGANLNGKISVTQGNVLLSGRPTPHARDFVNKSSAR
                     KDAHFSKNNEVVFEDDWINRTFKAAEITVNQSASFSSGRNVSDITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFGATQINGNVNLSQNAALVLGKAALW
                     GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKINH
                     LSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQDR
                     SRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQKA
                     TQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      983 a    684 c    639 g    686 t
ORIGIN
```

```
  Query Match             4.5%;  Score 197.8;  DB 1;  Length 2992;
  Best Local Similarity   57.9%;  Pred. No. 1e-33;
  Matches  371;  Conservative    0;  Mismatches  267;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          ||||  || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db   1035 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1094

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     || |||||| |     ||||||| ||| |||||
Db   1095 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCGGGG 1154

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |    || | ||  |  ||   |||| |||| ||      |     |||| |
Db   1155 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1214

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  |  | | || ||| |  |||||||
Db   1215 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1274

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||| || ||| || ||||  | || ||    ||    |||
Db   1275 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTC 1334

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||  | |||||  ||  | | |||
Db   1335 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1394

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| ||  | ||||||||| || || ||||| |||||||||   ||
Db   1395 AATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1454

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
           ||  ||     |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1455 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1514

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
           |||||   |  | | |  |||||  | ||  | || |||||   |   ||| (  |  |
Db   1515 CACAACACAAGCCACGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1574

Qy   1735 ggcaataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
             |||          |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1575 CAAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTAT 1634

Qy   1795 aaagatacgaccaaaacgaacgggcggctcaaccttgttta 1835
           |       |  ||| | |    ||    |||   | ||||
Db   1635 TACAGCTATAGAAAACCCATTCCGCAAGGCAAAGATCTTTA 1675


RESULT  21
AE002424/c
LOCUS       AE002424               10485 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 66 of 206 of
            the complete genome.
ACCESSION   AE002424 AE002098
VERSION     AE002424.1  GI:7225923
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10485)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10485)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10485
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            complement(67. .795)
                     /gene="NMB0696"
     CDS             complement(67. .795)
                     /gene="NMB0696"
                     /note="similar to GB:M61017 SP:P27675 PID:142988 percent
                     identity: 81.40; identified by sequence similarity;
                     putative"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="amino acid ABC transporter, ATP-binding protein"
                     /protein_id="AAF41113.1"
                     /db_xref="GI:7225924"
                     /translation="MIKFKNVHKHFKDLHVINGVNLEIKKGEVVVVCGPSGSGKSTLI
                     RTVNQLESIESGEIWVDGVNVADPKTDLNKIREEVGFVFQGFNLYPHLTVLENITLAP
                     MKVKGQNAEQAEKKAMELLERVGLAHKKDAFPSQLSGGQQQRVAIARGLAMEPRVMLF
                     DEPTSALDPEMVGEVLKVMKDLAESGMTMMCVTHEMGFAREVADRVIFVDKGQILEDE
                     TPEAFFTNPKHERAKQFLQQVMTH"
gene                 complement(968. .1747)
                     /gene="NMB0697"
CDS                  complement(968. .1747)
                     /gene="NMB0697"
                     /note="similar to GB:D10483 SP:P06992 GB:M11054 GB:M68521
                     GB:X04711 percent identity: 69.41; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dimethyladenosine transferase"
                     /protein_id="AAF41114.1"
                     /db_xref="GI:7225925"
                     /translation="MKEHKARKRFGQNFLQDTRIISDIVNAVRPQADDVVIEIGPGLA
                     AITEPLAKKLNRLHVVEIDRDIVCRLKTLPFADKLVIHEGDVLQFDFNGIAGKKKIVG
                     NLPYNISTPLLFKLAEVADDVVDMHFMLQKEVVERMVAAPKSNDYGRLGVMLQYFFDM
                     EMLIDVPPESFDPAPKVDSAVVRMIPVKHRIGKADDFEHFAKLVKLAFHQRRKTIRNN
                     LKELAGDDDLQAVGINPQDRAEHIAPEKYVALSNYLAGKVV"
gene                 1944. .2633
                     /gene="NMB0698"
CDS                  1944. .2633
                     /gene="NMB0698"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41115.1"
                     /db_xref="GI:7225926"
                     /translation="MLFEGLPCLAVWCKNRLSAVLPVGGLNVSENLLEIETHPFDPVL
                     PPKAAVMMMGTFPPKEDKRAMQFHYPNFQNDMWRVYGLVFFNDAAHFQRLSEKAFDAE
                     KIKAFLHERGIASCPTVLKAVRQHGNASDKFLKVVETVDLAAVLAKIPECRHICTTGG
                     KATEILLDIQGGGIKMPKTGETVPFPFAGRDLTLTRLPSTSRAYPLSLAKKAAAYRAF
                     FEMAGLCEKQL"
gene                 2716. .3918
                     /gene="NMB0699"
CDS                  2716. .3918
                     /gene="NMB0699"
                     /note="similar to GB:M58444 SP:P16706 PID:141798
                     PID:461373 percent identity: 80.71; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tryptophan synthase, beta subunit"
                     /protein_id="AAF41116.1"
                     /db_xref="GI:7225927"
                     /translation="MKNYHAPDEKGFFGEHGGLYVSETLIPALQELADAYKAAKNDPE
                     FWEAFRHDLKHYVGRPSPVYHAARLSEHLGGAQIWLKREDLNHTGAHKVNNTIGQALL
                     AKRMGKKRVIAETGAGQHGVASATVAARFGMTCDVYMGADDIQRQMPNVFRMKLLGAN
                     VVGVESGSRTLKDAMNEAMREWVARVDDTFYIIGTAAGPAPYPEMVRDFQCVIGNEAK
                     AQMQEAIGRQPDVAVACVGGGSNAIGLFHPYIGEENVRLVGVEAGGLGVNTPDHAAPI
                     TSGAPIGVLHGFRSYLMQDENGQVLGTHSVSAGLDYPGIGPEHSHLHDIKRVEYTVAK
                     DDEALEAFDLLCRFEGIIPALESSHAVAWAVKNAPKMGKDQVILVNLSGRGDKDINTV
                     AKLKGIKL"
gene                 complement(4097. .9544)
                     /gene="NMB0700"
CDS                  complement(4097. .9544)
                     /gene="NMB0700"
                     /note="similar to GP:732874 percent identity: 93.40;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA-specific serine endopeptidase"
                     /protein_id="AAF41117.1"
                     /db_xref="GI:7225928"
                     /translation="MKTKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHV
                     KGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFV
                     TEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGT
                     PYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDK
                     QEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHHWKTTGT
                     NSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGINN
                     DITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVIL
                     NQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEH
                     IRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYIQNNDDDDAGYYYYRPRKP
                     IPQGKDLYFKNYRYYALKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNN
                     QRISGFSGFFGEENGKGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSG
                     RPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVS
                     DITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNV
                     NLNQNAALVLGKAALWGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNA
                     SDAQSANKYHTLKINHLSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEP
                     NQEGLNLFDASSVQDRSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVK
                     PAPSPATNTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEAKRQQAKAEQVKRQ
                     QAEAERKSAELAKQKAEAEREARELATRQKAEQERSSAELARRHEKEREAAELSAKQK
                     VEAEREAQALAVRRKAEAEEAKRQAAELARRHEKEREAAELSAKQRVGEEERRQTAQS
                     QPQRRKRRAAPQDYMAASQDRPKRRGHRSVQQNNVEIAQAQAELARRQQEERKAAELL
                     AKQRAEAEREAQALAARRKAEAEEAKRQAAELAHRQEAERKAAELSANQKAAAEAQAL
                     AARQQKALARQQEEARKAAELAVKQKAETERKTAELAKQRAAAEAAKRQQEARQTAEL
                     ARRQEAERQAAELSAKQKAETDREAAESAKRKAEEEEHRQAAQSQPQRRKRRAAPQDY
                     MAASQNRPKRRGRRSTLPAPPSPSFDSSAYAAPRALHNPDWYENDYEEIPLDALEDEN
                     VSESVDTSDKQPQDNTELHEKYENDYEEIPLDALEDEDVSESVDTSDKQPQDNTELHE
                     KVETVSLQPRAAQPRAQAATQLQAQAAAQADAVSTNTNSALSDAMASTQSILLDTGAS
                     LTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRSLSENMQI
                     GGVLTYSDSQHTFDQASGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSRLQTQQKA
                     NFNRASIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGNDSVKVSSMSVKTLT
                     AGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGNSFVYKADNQQQYSAGAALLY
                     RNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
gene                 complement(9703. .9957)
                     /gene="NMB0701"
CDS                  complement(9703. .9957)
                     /gene="NMB0701"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41118.1"
                     /db_xref="GI:7225929"
                     /translation="MTNPSVRKPAPRHSHEPPSRHYHESGNLGLRVKKTYIPSFPQKQ
                     KTKIRNLKSRHSRKSGNPKTKSRRNLTEEPNPENRLPAFA"
BASE COUNT     2182 a   2461 c   2686 g   3156 t
ORIGIN


  Query Match             4.5%;  Score 196.2;  DB 1;  Length 10485;
  Best Local Similarity   57.7%;  Pred. No. 2.9e-33;
  Matches  370;  Conservative    0;  Mismatches  268;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |     |||||||| ||||| || || ||  | | |
Db   8464 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 8405

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     |  |||||| ||    ||||||| ||| |||||
Db   8404 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 8345

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
          || || |    || | ||  |  ||   |||| |||| ||       |     |||| |
Db   8344 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 8285

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
          || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   8284 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 8225

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
          | |  || |||||||||| || ||| || |||| || || || |  |    |||
Db   8224 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 8165
```

```
Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||| | |||||  ||  | | |||
Db   8164 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 8105

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   8104 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 8045

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    | | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   8044 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 7985

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   || | | |  |||||  | ||  | || || ||       |||  |     ||
Db   7984 CACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGATTACCGATCCC 7925

Qy   1735 ggcaataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
            | | |   |      |||  ||||  |   ||   || | | | | | | | | |
Db   7924 AAAACCATCTCTATTCATTATATTCAAAATAATGATGATGACGATGCTGGTTATTACTAT 7865

Qy   1795 aaagatacgaccaaaacgaacgggcggctcaaccttgttta 1835
            |    | |  ||| |||     |    |||   | ||||
Db   7864 TACCGCCCTAGGAAACCGATTCCACAAGGCAAAGATCTTTA 7824
```

RESULT 22
AX044030/c
LOCUS AX044030 349980 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 109 from Patent WO0066791.
ACCESSION AX044030
VERSION AX044030.1 GI:11342914
KEYWORDS .
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 349980)
AUTHORS Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C., Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M., Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
TITLE Neisseria genomic sequences and methods of their use
JOURNAL Patent: WO 0066791-A 109 09-NOV-2000;
CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES Location/Qualifiers
source 1. .349980
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
/note="sequence too long, cut in 8 pieces.-seq 1: 1 to 349980 349980 bases-seq 108: 300001 to 649980 349980 bases-seq 109: 600001 to 949980 349980 bases-seq 110: 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to 2272325 172325 bases"
BASE COUNT 82523 a 82940 c 96712 g 87805 t
ORIGIN

Query Match 4.5%; Score 196.2; DB 6; Length 349980;
Best Local Similarity 57.7%; Pred. No. 5.2e-33;
Matches 370; Conservative 0; Mismatches 268; Indels 3; Gaps 1;

```
Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |       |||||||| ||||| || || || | | |
Db 130038 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 129979

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     | |||||| ||    ||||||| ||| |||||
Db 129978 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 129919

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
            || || |    || | ||  |  ||   |||| |||| ||       |   |||| |
Db 129918 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 129859

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
          || | ||||||||||||||||||  |  |  | | | || ||| |  ||||||||
Db 129858 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 129799

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
          | |  || |||||||||||| || ||| || |||| || || || |   |   |||
Db 129798 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 129739

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||| | |||||  ||  | | |||
Db 129738 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 129679

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db 129678 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 129619

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    | | ||  | |  || |  ||    |||||||| |||   |||||||||
Db 129618 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 129559

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   || | | |  |||||  | ||  | || || ||       |||  |     ||
Db 129558 CACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGATTACCGATCCC 129499

Qy   1735 ggcaataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
            | | |   |      |||  ||||  |   ||   || | | | | | | | | |
Db 129498 AAAACCATCTCTATTCATTATATTCAAAATAATGATGATGACGATGCTGGTTATTACTAT 129439

Qy   1795 aaagatacgaccaaaacgaacgggcggctcaaccttgttta 1835
            |    | |  ||| |||     |    |||   | ||||
Db 129438 TACCGCCCTAGGAAACCGATTCCACAAGGCAAAGATCTTTA 129398
```

RESULT 23
NMIGA771
LOCUS NMIGA771 1489 bp DNA linear BCT 14-MAR-1995
DEFINITION N.meningitidis iga gene (EN3771).
ACCESSION X82468
VERSION X82468.1 GI:732865
KEYWORDS iga gene; IgA1 protease.
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 1489)
AUTHORS Lomholt,H., Poulsen,K. and Kilian,M.
TITLE Comparative characterization of the iga gene encoding IgA1 protease in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus influenzae
JOURNAL Mol. Microbiol. 15 (3), 495-506 (1995)
MEDLINE 95302961
REFERENCE 2 (bases 1 to 1489)
AUTHORS Lomholt,H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology, The Bartholin Building, University of Aarhus, 8000 C Aarhus C, DENMARK
FEATURES Location/Qualifiers
source 1. .1489
/organism="Neisseria meningitidis"
/strain="EN3771"
/db_xref="taxon:487"
gene 1. .1489
/gene="iga"
CDS <1. .>1489
/gene="iga"
/codon_start=1
/transl_table=11
/product="IgA1 protease"
/protein_id="CAA57851.1"
/db_xref="GI:732866"
/db_xref="SPTREMBL:Q57309"

/translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYT
VKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
DLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI"
BASE COUNT 492 a 321 c 325 g 351 t
ORIGIN

Query Match 4.5%; Score 195.6; DB 1; Length 1489;
Best Local Similarity 61.2%; Pred. No. 2.9e-33;
Matches 334; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

Qy 1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
Db 925 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 984

Qy 1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
Db 985 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGG 1044

Qy 1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
Db 1045 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1104

Qy 1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
Db 1105 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1164

Qy 1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
Db 1165 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1224

Qy 1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
Db 1225 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1284

Qy 1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
Db 1285 AATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1344

Qy 1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
Db 1345 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1404

Qy 1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
Db 1405 CACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1464

Qy 1735 ggcaat 1740
Db 1465 CAAAAT 1470

RESULT 24
NMIGATH2
LOCUS NMIGATH2 1489 bp DNA linear BCT 14-MAR-1995
DEFINITION N.meningitidis iga gene (ETH2).
ACCESSION X82469
VERSION X82469.1 GI:732887
KEYWORDS iga gene; IgA1 protease.
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 1489)
AUTHORS Lomholt,H., Poulsen,K. and Kilian,M.
TITLE Comparative characterization of the iga gene encoding IgA1 protease
in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
influenzae
JOURNAL Mol. Microbiol. 15 (3), 495-506 (1995)
MEDLINE 95302961
REFERENCE 2 (bases 1 to 1489)
AUTHORS Lomholt,H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
DENMARK
FEATURES Location/Qualifiers
source 1. .1489
/organism="Neisseria meningitidis"
/strain="ETH2"
/db_xref="taxon:487"
gene 1. .1489
/gene="iga"
CDS <1. .>1489
/gene="iga"
/codon_start=1
/transl_table=11
/product="IgA1 protease"
/protein_id="CAA57852.1"
/db_xref="GI:732888"
/db_xref="SPTREMBL:Q57309"
/translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYT
VKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
DLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI"
BASE COUNT 492 a 321 c 325 g 351 t
ORIGIN

Query Match 4.5%; Score 195.6; DB 1; Length 1489;
Best Local Similarity 61.2%; Pred. No. 2.9e-33;
Matches 334; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

Qy 1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
Db 925 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 984

Qy 1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
Db 985 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGG 1044

Qy 1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
Db 1045 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1104

Qy 1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
Db 1105 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1164

Qy 1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
Db 1165 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1224

Qy 1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
Db 1225 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1284

Qy 1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
Db 1285 AATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1344

Qy 1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
Db 1345 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1404

```
Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   ||||   |  |||||  | ||  | || |||||   |   ||| |  |  |
Db   1405 CACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1464

Qy   1735 ggcaat 1740
             |||
Db   1465 CAAAAT 1470
```

RESULT 25
NMIGAG117
LOCUS NMIGAG117 1525 bp DNA linear BCT 14-MAR-1995
DEFINITION N.meningitidis iga gene (NG117).
ACCESSION X82483
VERSION X82483.1 GI:732881
KEYWORDS iga gene; IgA1 protease.
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 1525)
AUTHORS Lomholt,H., Poulsen,K. and Kilian,M.
TITLE Comparative characterization of the iga gene encoding IgA1 protease in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus influenzae
JOURNAL Mol. Microbiol. 15 (3), 495-506 (1995)
MEDLINE 95302961
REFERENCE 2 (bases 1 to 1525)
AUTHORS Lomholt,H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology, The Bartholin Building, University of Aarhus, 8000 C Aarhus C, DENMARK
FEATURES Location/Qualifiers
source 1. .1525
/organism="Neisseria meningitidis"
/strain="NG117"
/db_xref="taxon:487"
gene 1. .1525
/gene="iga"
CDS <1. .>1525
/gene="iga"
/codon_start=1
/transl_table=11
/product="IgA1 protease"
/protein_id="CAA57866.1"
/db_xref="GI:732882"
/db_xref="SPTREMBL:Q51171"
/translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYA
VSVKHAKAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNM
ARFNKFVTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKG
YDLRDLSQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALT
NYGVLGDSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDN
AGTIKGNREHHWNITFGTNSKIGSTAVRLANNERDANNGQNVTFEDNGTLVLDQNINQ
GAGGLFFKGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEIN
GTGVNQGQLKVGDGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYF
GFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI
"
BASE COUNT 507 a 318 c 334 g 366 t
ORIGIN

Query Match 4.5%; Score 195.6; DB 1; Length 1525;
Best Local Similarity 61.2%; Pred. No. 2.9e-33;
Matches 334; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

```
Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          ||||  || |||   ||| || |       ||||||||| ||||| || || ||  | | |
Db    961 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1020

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     || |||||| ||    ||||||| ||| |||||
Db   1021 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGG 1080

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
          || || |    || | ||  |  ||   |||| |||| ||      |     |||| |
Db   1081 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1140

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
          || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1141 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1200

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
          | |  || |||||||||||| || ||| || |||| || || || |   |    |||
Db   1201 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1260

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||| | |||||  ||  | | |||
Db   1261 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1320

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| ||  | |||||||| || || ||||| ||||||||   ||
Db   1321 AATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1380

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    | | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1381 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1440

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   ||||   |  |||||  | ||  | || |||||   |   ||| |  |  |
Db   1441 CACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1500

Qy   1735 ggcaat 1740
             |||
Db   1501 CAAAAT 1506
```

RESULT 26
NGIGAPRO
LOCUS NGIGAPRO 4899 bp DNA linear BCT 21-MAR-1995
DEFINITION Neisseria gonorrhoeae gene for IgA protease.
ACCESSION X04835
VERSION X04835.1 GI:44868
KEYWORDS IgA protease; inverted repeat; protease; secreted protein; signal peptide.
SOURCE Neisseria gonorrhoeae.
ORGANISM Neisseria gonorrhoeae
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 4899)
AUTHORS Pohlner,J., Halter,R., Beyreuther,K. and Meyer,T.F.
TITLE Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA protease
JOURNAL Nature 325 (6103), 458-462 (1987)
MEDLINE 87115823
FEATURES Location/Qualifiers
source 1. .4899
/organism="Neisseria gonorrhoeae"
/strain="M11"
/db_xref="taxon:485"
/clone="pIP100"
promoter 34. .39
/note="put. -43 region"
promoter 43. .48
/note="put. -35 region"
promoter 63. .68
/note="put. -10 region"
RBS 87. .97
/note="put. rRNA binding region"
CDS 104. .4702
/note="IgA protease precursor; 169K (AA -27 to 1505)"
/codon_start=1
/transl_table=11
/protein_id="CAA28538.1"
/db_xref="GI:44869"
/db_xref="SWISS-PROT:P09790"

```
                     /translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
                     KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
                     VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
                     SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
                     DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
                     NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
                     KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
                     QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
                     DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
                     DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
                     ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQKRFLLTGGANLNGKI
                     SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
                     SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
                     SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
                     LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
                     QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
                     NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
                     QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
                     ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
                     IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
                     STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
                     LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
                     LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
                     LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
                     MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
                     AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
     sig_peptide     104. .184
     mat_peptide     185. .4699
                     /product="mature IgA protease (AA 1-1505)"
     misc_feature    3061. .3062
                     /note="autoproteolytic site for IgA protease"
     misc_feature    3157. .3158
                     /note="autoproteolytic site for IgA protease"
     misc_feature    3466. .3467
                     /note="autoproteolytic site for IgA protease"
     repeat_unit     4725. .4739
                     /note="inverted repeat A"
     terminator      4725. .4762
                     /note="put. transcription termination signal"
     repeat_unit     4747. .4762
                     /note="inverted repeat A'"
BASE COUNT     1617 a   1104 c   1080 g   1098 t
ORIGIN


  Query Match              4.5%;  Score 195.6;  DB 1;  Length 4899;
  Best Local Similarity   62.3%;  Pred. No. 3.5e-33;
  Matches  344;  Conservative    0;  Mismatches  199;  Indels    9;  Gaps    2;

Qy   1167 gaataatggagaaaatatttcctttattgacgaaggaaaaggcgaattgatacttaccag 1226
          ||| || ||| ||||| || |||||      ||||  || |||   ||| || | |
Db   1192 GAACAACGGACAAAATGTTACCTTT------GAAGACAACGGCACTTTGGTATTGAATCA 1245

Qy   1227 caacatcaatcaaggtgctggaggattatatttccaaggagattttacggtctcgcctga 1286
           |||||||| ||||| || || ||  | | |||| |||| ||||  || |||      |
Db   1246 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 1305

Qy   1287 aaataacgaa---acttggcaaggcgcgggcgttcatatcagtgaagacagtaccgttac 1343
          ||||||  ||    ||||||  ||| |||||  || || |    || | ||  |  ||
Db   1306 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 1365

Qy   1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgca 1403
          |||| |||| ||      |     |||| |  || | ||||||||||||||||  || |
Db   1366 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 1425

Qy   1404 cgttcaagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtacagtcatttt 1463
            |  | | || ||| |  ||||||||     | |  || ||||||||||| || ||| |
Db   1426 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 1485

Qy   1464 ggatcagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523
          | ||||  | || || |   |    |||   || || ||     || ||||| | |||||
Db   1486 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 1545

Qy   1524 cggcaggggtacggtgcaactgaatgccgataatcagttcaaccccgacaaaactctattt 1583
          |||  | |||||  ||  | | |||      ||||||  | || ||||| || || |||||
Db   1546 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 1605

Qy   1584 cggctttcgcggcggacgtttggatttaaacgggcattcgctttcgttccaccgtattca 1643
          ||| || || ||||| |||||||||   || ||  ||    |  | ||  | |  || |
Db   1606 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 1665

Qy   1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703
           ||    |||||||| |||   |||||||||||||||   |||| | |  |||||  | ||
Db   1666 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 1725

Qy   1704 cattacaggcaa 1715
           | || || ||
Db   1726 ACTAACGGGTAA 1737


RESULT  27
A02796
LOCUS       A02796                  4899 bp    DNA     linear   PAT 10-MAY-1993
DEFINITION  N.gonorrhoeae gene for IgA-protease precursor.
ACCESSION   A02796
VERSION     A02796.1  GI:345152
KEYWORDS    IgA protease.
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   Meyer,T.F., Halter,R. and Pohlner,J.
  TITLE     Process for recovering proteins by means of gene technology using
            gram-negative host cells
  JOURNAL   Patent: EP 0254090-A 1 27-JAN-1988;
            Max-Planck-Gesellschaft zur Foerderung der Wissenschaften
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
     CDS             104. .4702
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA00270.1"
                     /db_xref="GI:345153"
                     /translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
                     KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
                     VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
                     SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
                     DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
                     NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
                     KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
                     QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
                     DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
                     DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
                     ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQNRFLLTGGANLNGKI
                     SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
                     SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
                     SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
                     LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
                     QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
                     NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
                     QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
                     ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
                     IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
                     STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
                     LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
                     LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
                     LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
                     MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
                     AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
```

```
BASE COUNT     1617 a   1105 c   1079 g   1098 t
ORIGIN

  Query Match             4.5%;  Score 195.6;  DB 6;  Length 4899;
  Best Local Similarity   62.3%;  Pred. No. 3.5e-33;
  Matches  344;  Conservative    0;  Mismatches  199;  Indels    9;  Gaps    2;

Qy    1167 gaataatggagaaaatatttccttttattgacgaaggaaaaggcgaattgatacttaccag 1226
           ||| || ||| ||||| || |||||      ||||  || |||   ||| || | |
Db    1192 GAACAACGGACAAAATGTTACCTTT------GAAGACAACGGCACTTTGGTATTGAATCA 1245

Qy    1227 caacatcaatcaaggtgctggaggattatatttccaaggagattttacggtctcgcctga 1286
           |||||||| ||||| || || ||  | | |||| |||| ||||  || |||    .  |
Db    1246 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 1305

Qy    1287 aaataacgaa---acttggcaaggcgcgggcgttcatatcagtgaagacagtaccgttac 1343
           |||||| ||    ||||||  ||| ||||| || || |    || | ||  |  ||
Db    1306 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 1365

Qy    1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgca 1403
           |||| |||| ||      |     |||| |  || | ||||||||||||||||||  || |
Db    1366 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 1425

Qy    1404 cgttcaagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtacagtcatttt 1463
             |  | | || ||| |  |||||||||     | |  || ||||||||||| || ||| |
Db    1426 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 1485

Qy    1464 ggatcagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523
           | ||||  | || || |   |    |||  || || ||     || ||||| | |||||
Db    1486 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 1545

Qy    1524 cggcaggggtacggtgcaactgaatgccgataatcagttcaaccccgacaaactctattt 1583
           |||  | |||||  ||  | | |||      |||||| | || ||||| || || |||||
Db    1546 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 1605

Qy    1584 cggctttcgcggcggacgtttggatttaaacgggcattcgctttcgttccaccgtattca 1643
           ||| || || ||||| |||||||||   || ||  ||    |  | ||  | |  || |
Db    1606 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 1665

Qy    1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703
             ||    |||||||| |||   ||||||||||||||   |||| | |  |||||  | ||
Db    1666 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 1725

Qy    1704 cattacaggcaa 1715
             | || || ||
Db    1726 ACTAACGGGTAA 1737


RESULT  28
A12416
LOCUS       A12416                   4899 bp    DNA     linear   PAT 05-OCT-1994
DEFINITION  IgA-protease gene precursor.
ACCESSION   A12416
VERSION     A12416.1  GI:640585
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   .
  TITLE     PROCESS FOR THE GENETIC PRODUCTION OF PROTEINS BY USING OF
            GRAM-NEGATIVE HOST CELLS
  JOURNAL   Patent: WO 8800238-A 12 14-JAN-1988;
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="Neisseria gonorrhoeae"
                     /db_xref="taxon:485"
     gene            104. .4702
                     /gene="IgA protease precursor"
     CDS             104. .4702
                     /gene="IgA protease precursor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA01030.1"
                     /db_xref="GI:640586"
                     /translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
                     KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
                     VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
                     SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
                     DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
                     NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
                     KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
                     QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
                     DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
                     DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
                     ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQNRFLLTGGANLNGKI
                     SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
                     SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
                     SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
                     LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
                     QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
                     NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
                     QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
                     ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
                     IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
                     STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
                     LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
                     LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
                     LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
                     MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
                     AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
BASE COUNT     1617 a   1105 c   1079 g   1098 t
ORIGIN

  Query Match             4.5%;  Score 195.6;  DB 6;  Length 4899;
  Best Local Similarity   62.3%;  Pred. No. 3.5e-33;
  Matches  344;  Conservative    0;  Mismatches  199;  Indels    9;  Gaps    2;

Qy    1167 gaataatggagaaaatatttccttttattgacgaaggaaaaggcgaattgatacttaccag 1226
           ||| || ||| ||||| || |||||      ||||  || |||   ||| || | |
Db    1192 GAACAACGGACAAAATGTTACCTTT------GAAGACAACGGCACTTTGGTATTGAATCA 1245

Qy    1227 caacatcaatcaaggtgctggaggattatatttccaaggagattttacggtctcgcctga 1286
           |||||||| ||||| || || ||  | | |||| |||| ||||  || |||      |
Db    1246 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 1305

Qy    1287 aaataacgaa---acttggcaaggcgcgggcgttcatatcagtgaagacagtaccgttac 1343
           |||||| ||    ||||||  ||| ||||| || || |    || | ||  |  ||
Db    1306 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 1365

Qy    1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgca 1403
           |||| |||| ||      |     |||| |  || | ||||||||||||||||||  || |
Db    1366 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 1425

Qy    1404 cgttcaagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtacagtcatttt 1463
             |  | | || ||| |  |||||||||     | |  || ||||||||||| || ||| |
Db    1426 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 1485

Qy    1464 ggatcagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523
           | ||||  | || || |   |    |||  || || ||     || ||||| | |||||
Db    1486 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 1545

Qy    1524 cggcaggggtacggtgcaactgaatgccgataatcagttcaaccccgacaaactctattt 1583
           |||  | |||||  ||  | | |||      |||||| | || ||||| || || |||||
Db    1546 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 1605

Qy    1584 cggctttcgcggcggacgtttggatttaaacgggcattcgctttcgttccaccgtattca 1643
           ||| || || ||||| |||||||||   || ||  ||    |  | ||  | |  || |
Db    1606 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 1665

Qy    1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703
             ||    |||||||| |||   ||||||||||||||   |||| | |  |||||  | ||
```

```
Db   1666 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 1725

Qy   1704 cattacaggcaa 1715
          | || || ||
Db   1726 ACTAACGGGTAA 1737


RESULT 29
AR079202
LOCUS       AR079202                4899 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 5 from patent US 5965424.
ACCESSION   AR079202
VERSION     AR079202.1  GI:10005948
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   Ambrosius,D., Dony,C. and Rudolph,R.
  TITLE     Methods for making neisseria or hemophilus IgA protease and DNA
            encoding the proteases
  JOURNAL   Patent: US 5965424-A 5 12-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="unknown"
BASE COUNT     1617 a   1104 c   1080 g   1098 t
ORIGIN


  Query Match             4.5%;  Score 195.6;  DB 6;  Length 4899;
  Best Local Similarity   62.3%;  Pred. No. 3.5e-33;
  Matches  344;  Conservative    0;  Mismatches  199;  Indels    9;  Gaps    2;

Qy   1167 gaataatggagaaaatatttcctttattgacgaaggaaaaggcgaattgatacttaccag 1226
          ||| || ||| ||||| || |||||      ||||  || |||   ||| || | |
Db   1192 GAACAACGGACAAAATGTTACCTTT------GAAGACAACGGCACTTTGGTATTGAATCA 1245

Qy   1227 caacatcaatcaaggtgctggaggattatatttccaaggagattttacggtctcgcctga 1286
           |||||||| ||||| || || ||  | | |||| |||| |||| || |||      |
Db   1246 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 1305

Qy   1287 aaataacgaa---acttggcaaggcgcgggcgttcatatcagtgaagacagtaccgttac 1343
          |||||| ||    ||||||  ||| |||||  || || |    || | ||  |  ||
Db   1306 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 1365

Qy   1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgca 1403
          |||| |||| ||       |     |||| |  || | ||||||||||||||||  || |
Db   1366 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 1425

Qy   1404 cgttcaagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtacagtcatttt 1463
            |  | | || ||| |  |||||||      | |  || ||||||||||| || ||| |
Db   1426 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 1485

Qy   1464 ggatcagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523
          | ||||  | || || |    |     |||   || || ||     || ||||| | |||||
Db   1486 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 1545

Qy   1524 cggcaggggtacggtgcaactgaatgccgataatcagttcaaccccgacaaaactctattt 1583
          |||  | |||||  ||  | | |||      ||||| | || ||||| || || |||||
Db   1546 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 1605

Qy   1584 cggctttcgcggcggacgtttggatttaaacgggcattcgctttcgttccaccgtattca 1643
          ||| || || ||||| |||||||||   || ||  ||    |  | ||  | |  || |
Db   1606 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 1665

Qy   1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703
           ||    |||||||| |||   |||||||||||||||   |||| | |  |||||  | ||
Db   1666 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 1725

Qy   1704 cattacaggcaa 1715
          | || || ||
Db   1726 ACTAACGGGTAA 1737
```

```
RESULT 30
NGCOMLGN
LOCUS       NGCOMLGN                8393 bp    DNA     linear   BCT 18-NOV-1996
DEFINITION  N.gonorrhoeae gene for competence lipoprotein ComL.
ACCESSION   Z49895
VERSION     Z49895.1  GI:1107832
KEYWORDS    ComL; competence lipoprotein.
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 8393)
  AUTHORS   Fussenegger,M., Facius,D., Meier,J. and Meyer,T.F.
  TITLE     A novel peptidoglycan-linked lipoprotein (ComL) that functions in
            natural transformation competence of Neisseria gonorrhoeae
  JOURNAL   Mol. Microbiol. 19 (5), 1095-1105 (1996)
  MEDLINE   96249702
REFERENCE   2  (bases 1 to 8393)
  AUTHORS   Fussenegger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1995) Martin Fussenegger Dr., Infektionsbiologie,
            Max-Planck-Institut, fuer Biologie, Spemannstr. 34, Tuebingen, BW,
            D-72076, Germany
FEATURES             Location/Qualifiers
     source          1. .8393
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11A"
                     /db_xref="taxon:485"
     CDS             120. .923
                     /codon_start=1
                     /transl_table=11
                     /product="ComL, competence lipoprotein"
                     /protein_id="CAA90076.1"
                     /db_xref="GI:1107833"
                     /db_xref="SPTREMBL:Q50985"
                     /translation="MKKILLTVSLGLALSACATQGTADKDAQITQDWSVEKLYAEAQD
                     ELNSSNYTRAVKLYEILESRFPTSRHARQSQLDTAYAYYKDDEKDKALAAIERFRRLH
                     PQHPNMDYALYLRGLVLFNEDQSFLNKLASQDWSDRDPKANREAYQAFAELVQRFPNS
                     KYAADATARMVKLVDALGGNEMSVARYYMKRGAYIAAANRAKKIIGSYQNTRYVEESL
                     AILELAYKKLDKPQLAADTRRVLETNFPKSPFLTHAWQPDDMPWWRYWH"
     terminator      934. .967
BASE COUNT     2550 a   2300 c   1886 g   1657 t
ORIGIN


  Query Match             4.5%;  Score 195.6;  DB 1;  Length 8393;
  Best Local Similarity   62.3%;  Pred. No. 3.8e-33;
  Matches  344;  Conservative    0;  Mismatches  199;  Indels    9;  Gaps    2;

Qy   1167 gaataatggagaaaatatttcctttattgacgaaggaaaaggcgaattgatacttaccag 1226
          ||| || ||| ||||| || |||||      ||||  || |||   ||| || | |
Db   4686 GAACAACGGACAAAATGTTACCTTT------GAAGACAACGGCACTTTGGTATTGAATCA 4739

Qy   1227 caacatcaatcaaggtgctggaggattatatttccaaggagattttacggtctcgcctga 1286
           |||||||| ||||| || || ||  | | |||| |||| |||| || |||      |
Db   4740 AAACATCAACCAAGGCGCGGGCGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGC 4799

Qy   1287 aaataacgaa---acttggcaaggcgcgggcgttcatatcagtgaagacagtaccgttac 1343
          ||||||  ||    ||||||  ||| |||||  || || |    || | ||  |  ||
Db   4800 AAATAATGACATCACTTGGTTAGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGT 4859

Qy   1344 ttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgca 1403
          |||| |||| ||       |     |||| |  || | ||||||||||||||||  || |
Db   4860 TTGGCAAGTCAAAAATCCGAATGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGA 4919

Qy   1404 cgttcaagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtacagtcatttt 1463
            |  | | || ||| |  |||||||      | |  || ||||||||||| || ||| |
Db   4920 AATAAACGGCACAGGCGTTAACCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCT 4979

Qy   1464 ggatcagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcggcttggtcag 1523
          | ||||  | || || |    |     |||   || || ||     || ||||| | |||||
Db   4980 GAATCAAAAAGCCGATGCCGACAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAG 5039
```

```
Qy   1524 cggcaggggtacggtgcaactgaatgccgataatcagttcaaccccgacaaactctattt 1583
          |||  | |||||  ||  | | |||      ||||||| | || ||||| || || |||||
Db   5040 CGGACGCGGTACATTGGTATTAAATAGTTCAAATCAGATTAATCCCGATAACCTATATTT 5099

Qy   1584 cggctttcgcggcggacgtttggatttaaacgggcattcgctttcgttccaccgtattca 1643
          ||| || || ||||| |||||||||   || ||  ||    |  | ||  | |  || |
Db   5100 CGGTTTCCGTGGCGGTCGTTTGGATGCCAATGGCAATGACTTGACTTTTGAACACATCCG 5159

Qy   1644 aaataccgatgaaggggcgatgattgtcaaccacaatcaagacaaagaatccaccgttac 1703
            ||    |||||||| |||   |||||||||||||   |||| | |  |||||  | ||
Db   5160 CAACGTGGATGAAGGCGCGCGCATTGTCAACCACAACACAGACCACGCCTCCACAATCAC 5219

Qy   1704 cattacaggcaa 1715
             | || ||  ||
Db   5220 ACTAACGGGTAA 5231
```

RESULT 31
NMIGA476
LOCUS NMIGA476 1492 bp DNA linear BCT 14-MAR-1995
DEFINITION N.meningitidis iga gene (NG44/76).
ACCESSION X82481
VERSION X82481.1 GI:732863
KEYWORDS iga gene; IgA1 protease.
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 1492)
AUTHORS Lomholt,H., Poulsen,K. and Kilian,M.
TITLE Comparative characterization of the iga gene encoding IgA1 protease in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus influenzae
JOURNAL Mol. Microbiol. 15 (3), 495-506 (1995)
MEDLINE 95302961
REFERENCE 2 (bases 1 to 1492)
AUTHORS Lomholt,H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology, The Bartholin Building, University of Aarhus, 8000 C Aarhus C, DENMARK

FEATURES Location/Qualifiers
source 1. .1492
/organism="Neisseria meningitidis"
/strain="NG44/76"
/db_xref="taxon:487"
gene 1. .1489
/gene="iga"
CDS <1. .>1489
/gene="iga"
/codon_start=1
/transl_table=11
/product="IgA1 protease"
/protein_id="CAA57864.1"
/db_xref="GI:732864"
/db_xref="SPTREMBL:Q51165"
/translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHH
WNITFGTNSKIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDY
TVKGINNDITWLGAGIDVTDGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKV
GDGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANG
NDLTFEHIRNVDEGARIVNHNTSHASTITLTGKSLITNPNSLSVHS"
BASE COUNT 500 a 317 c 322 g 353 t
ORIGIN

Query Match 4.4%; Score 193; DB 1; Length 1492;
Best Local Similarity 62.0%; Pred. No. 1.1e-32;
Matches 323; Conservative 0; Mismatches 195; Indels 3; Gaps 1;

```
Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          ||||  || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db    928 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 987

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     |  |||||| ||    ||||||| ||| |||||
Db    988 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 1047

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
          || || | |  || | ||  |  ||   |||| |||| ||      |    |||| |
Db   1048 ATTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1107

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
          || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1108 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1167

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
          | |  || |||||||||| || ||| || |||| || || || |   |    |||
Db   1168 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1227

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||| | ||||  ||  | | |||
Db   1228 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1287

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          ·|||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   1288 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1347

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1348 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1407

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
          |||||   |  | | |  |||||  | ||  | || || ||
Db   1408 CACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAA 1448
```

RESULT 32
AF012203
LOCUS AF012203 2980 bp DNA linear BCT 30-JAN-2001
DEFINITION Neisseria meningitidis strain Z2491 IgA1 protease precursor (iga) gene, partial cds.
ACCESSION AF012203
VERSION AF012203.1 GI:2529489
KEYWORDS .
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 2980)
AUTHORS Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del Valle,J. and Achtman,M.
TITLE Clonal descent and microevolution of Neisseria meningitidis during 30 years of epidemic spread
JOURNAL Mol. Microbiol. 25 (6), 1047-1064 (1997)
MEDLINE 98010345
REFERENCE 2 (bases 1 to 2980)
AUTHORS Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del Valle,J. and Achtman,M.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare Genetik, Ihnestrasse 73, Berlin 14195, Germany

FEATURES Location/Qualifiers
source 1. .2980
/organism="Neisseria meningitidis"
/strain="Z2491"
/db_xref="taxon:487"
/note="serogroup A, subgroup IV-1; isolated in Gambia, 1983"
mat_peptide 1. .>2980
/gene="iga"
/product="IgA1 protease"

```
     gene            <1. .>2980
                     /gene="iga"
                     /allele="iga4"
     CDS             <1. .>2980
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45786.1"
                     /db_xref="GI:2529490"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     GANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
                     GIVSGRGTLVLNSPDQINPNNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
                     RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAQ
                     KDAHFSKNNEVVFEDDWINRTFKATEITVNQSASFSSGRNVSNITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSPGATKINGNVNLNQNAALVLGKAALW
                     GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKINH
                     LSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVRDR
                     SHLSVSLANHHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQTD
                     SAQIAKPQNIVVAPPSPQAN"
BASE COUNT      974 a    687 c    642 g    677 t
ORIGIN
```

Query Match 4.4%; Score 193; DB 1; Length 2980;
Best Local Similarity 57.4%; Pred. No. 1.2e-32;
Matches 368; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

```
Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db   1035 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1094

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     || |||||| |     ||||||| ||| || ||
Db   1095 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCAGGG 1154

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |    || | ||  |  ||   |||| |||| ||      |     |||| |
Db   1155 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1214

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1215 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1274

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||| || ||| || |||| || || || |   |    |||
Db   1275 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1334

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | || |||||  | |||||  ||   |||||
Db   1335 CAGGCTTTCTCCCAAGTCGGCATTGTGAGCGGGCGCGGTACATTGGTGCTGAACAGTCCG 1394

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
           | ||| | |||||  | || || |||||||| || || ||||| |||||||||   ||
Db   1395 GACCAGATTAACCCGAATAATCTGTATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1454

Qy   1615 gggcattcgctttcgttccaccgtattcaaaatacgatgaaggggcgatgattgtcaac 1674
          ||  ||     |  | ||  | |  || |  ||    |||||||| |||   ||||||||||
Db   1455 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1514

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   ||||   |  ||||| | ||  | || ||||   |   ||| |  |  |
Db   1515 CACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1574

Qy   1735 ggcaataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
```

```
          |||         |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1575 CAAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTAT 1634

Qy   1795 aaagatacgaccaaaacgaacgggcggctcaaccttgttta 1835
          |       |  ||| | |    ||    |||   | ||||
Db   1635 TACAGCTATAGAAAACCCATTCCGCAAGGCAAAGATCTTTA 1675
```

```
RESULT  33
AF012204
LOCUS       AF012204               2980 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z3906 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012204
VERSION     AF012204.1  GI:2529491
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2980)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2980)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2980
                     /organism="Neisseria meningitidis"
                     /strain="Z3906"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup III; isolated in China, 1966;
                     also present in strain B1, serogroup A, subgroup IV-2
                     isolated in USA, 1937"
     mat_peptide     1. .>2980
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2980
                     /gene="iga"
                     /allele="iga2"
     CDS             <1. .>2980
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45787.1"
                     /db_xref="GI:2529492"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     GANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
                     GIVSGRGTLVLNSPDQINPNNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
                     RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAQ
                     KDAHFSKNNEVVFEDDWINRTFKATEITVNQSASFSSGRNVSNITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSPGATQINGNVNLNQNAALVLGKAALW
                     GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKINH
                     LSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVRDR
                     SHLSVSLANHHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQTD
                     SAQIAKPQNIVVAPPSPQAN"
BASE COUNT      973 a    688 c    642 g    677 t
```

ORIGIN

Query Match 4.4%; Score 193; DB 1; Length 2980;
Best Local Similarity 57.4%; Pred. No. 1.2e-32;
Matches 368; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

```
Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db   1035 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1094

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     || |||||| |     ||||||| ||| || ||
Db   1095 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCAGGG 1154

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |    || | ||  |  ||   |||| |||| ||       |     |||| |
Db   1155 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1214

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1215 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1274

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||||| || ||| || |||| || || || |   |    |||
Db   1275 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1334

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | || |||||  | |||||  ||   |||||
Db   1335 CAGGCTTTCTCCCAAGTCGGCATTGTGAGCGGGCGCGGTACATTGGTGCTGAACAGTCCG 1394

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
           | ||| | |||||  | || || |||||||| || || ||||| |||||||||   ||
Db   1395 GACCAGATTAACCCGAATAATCTGTATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1454

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1455 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1514

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   ||||   |  ||||| | ||  | || |||||   |   ||| |  |  |
Db   1515 CACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1574

Qy   1735 ggcaataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
           |||         |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1575 CAAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTAT 1634

Qy   1795 aaagatacgaccaaaacgaacgggcggctcaaccttgttta 1835
           |      |  ||| | |    ||    |||   | ||||
Db   1635 TACAGCTATAGAAAACCCATTCCGCAAGGCAAAGATCTTTA 1675
```

RESULT 34
AF012205
LOCUS AF012205 2980 bp DNA linear BCT 30-JAN-2001
DEFINITION Neisseria meningitidis strain Z3910 IgA1 protease precursor (iga) gene, partial cds.
ACCESSION AF012205
VERSION AF012205.1 GI:2529493
KEYWORDS .
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 2980)
AUTHORS Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del Valle,J. and Achtman,M.
TITLE Clonal descent and microevolution of Neisseria meningitidis during 30 years of epidemic spread
JOURNAL Mol. Microbiol. 25 (6), 1047-1064 (1997)
MEDLINE 98010345
REFERENCE 2 (bases 1 to 2980)
AUTHORS Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del Valle,J. and Achtman,M.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare Genetik, Ihnestrasse 73, Berlin 14195, Germany

FEATURES Location/Qualifiers
source 1. .2980
/organism="Neisseria meningitidis"
/strain="Z3910"
/db_xref="taxon:487"
/note="serogroup A, subgroup III; isolated in China, 1966"
mat_peptide 1. .>2980
/gene="iga"
/product="IgA1 protease"
gene <1. .>2980
/gene="iga"
/allele="iga1"
CDS <1. .>2980
/gene="iga"
/codon_start=3
/transl_table=11
/product="IgA1 protease precursor"
/protein_id="AAC45788.1"
/db_xref="GI:2529494"
/translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
IGFGDNSKHHSLEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
GANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
GIVSGRGTLVLNSPDQINPNNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
LKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAQ
KDAHFSKNNEVVFEDDWINRTFKATEITVNQSASFSSGRNVSNITANITATDNAKVNL
GYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAALW
GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKINH
LSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVRDR
SHLSVSLANHHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQTD
SAQIAKPQNIVVAPPSPQAN"
BASE COUNT 973 a 687 c 642 g 678 t
ORIGIN

Query Match 4.4%; Score 193; DB 1; Length 2980;
Best Local Similarity 57.4%; Pred. No. 1.2e-32;
Matches 368; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

```
Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db   1035 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1094

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     || |||||| |     ||||||| ||| || ||
Db   1095 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCAGGG 1154

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |    || | ||  |  ||   |||| |||| ||       |     |||| |
Db   1155 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1214

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1215 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1274

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||||| || ||| || |||| || || || |   |    |||
Db   1275 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1334

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | || |||||  | |||||  ||   |||||
Db   1335 CAGGCTTTCTCCCAAGTCGGCATTGTGAGCGGGCGCGGTACATTGGTGCTGAACAGTCCG 1394
```

```
Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          | ||| | |||||  | || || |||||||| || || ||||| |||||||||   ||
Db   1395 GACCAGATTAACCCGAATAATCTGTATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1454

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   ||||||||||
Db   1455 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1514

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   ||||   |  |||||  | ||  | || |||||   |   ||| |  |  |
Db   1515 CACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1574

Qy   1735 ggcaataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
             |||         |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db   1575 CAAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTAT 1634

Qy   1795 aaagatacgaccaaaacgaacgggcggctcaaccttgttta 1835
           |       |  ||| | |    ||    |||   | ||||
Db   1635 TACAGCTATAGAAAACCCATTCCGCAAGGCAAAGATCTTTA 1675


RESULT 35
AF012208
LOCUS       AF012208               2983 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4400 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012208
VERSION     AF012208.1  GI:2529499
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2983)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2983)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2983
                     /organism="Neisseria meningitidis"
                     /strain="Z4400"
                     /db_xref="taxon:487"
                     /note="serogroup C, ET-37 complex; isolated in Mali, 1990"
     mat_peptide     1. .>2983
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2983
                     /gene="iga"
                     /allele="iga5"
     CDS             <1. .>2983
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45791.2"
                     /db_xref="GI:12597968"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHHWNITFGTNSKIGSTAVRLAGNE
                     KDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGINNDITWLGAGIDVTDGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     SHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     RKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPVPSPATNTASQAQT
                     DSAQIAKPQNIVVAPPSPQAN"
BASE COUNT      978 a    679 c    639 g    687 t
ORIGIN


  Query Match             4.4%;  Score 193;  DB 1;  Length 2983;
  Best Local Similarity   62.0%;  Pred. No. 1.2e-32;
  Matches  323;  Conservative    0;  Mismatches  195;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          ||||  || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db   1038 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1097

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     | |||||| ||    ||||||| ||| |||||
Db   1098 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 1157

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || | |  || | ||  |  ||   |||| |||| ||      |    |||| |
Db   1158 ATTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1217

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  | | || ||| |  |||||||||
Db   1218 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1277

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||||| || ||| || |||| || || || |   |    |||
Db   1278 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1337

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||  | |||||  ||  | | |||
Db   1338 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1397

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   1398 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1457

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   ||||||||||
Db   1458 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1517

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
          |||||   |  | | |  |||||  | ||  | || || ||
Db   1518 CACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAA 1558


RESULT 36
AF012209
LOCUS       AF012209               2995 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4081 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012209
VERSION     AF012209.1  GI:2529501
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
```

```
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Neisseria meningitidis"
                     /strain="Z4081"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup V; isolated in China, 1979;
                     also present in strain Z4063, serogroup A, subgroup VII,
                     isolated in China, 1979"
     mat_peptide     1. .>2995
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2995
                     /gene="iga"
                     /allele="iga9"
     CDS             <1. .>2995
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45792.2"
                     /db_xref="GI:12597969"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHHWNITFGTNSKIGSTAVRLAGNE
                     KDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGINNDITWLGAGIDVTDGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     SHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     RKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASCHYQLHVQDKTGEPNQEGLNLFDASSVQD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQK
                     ATQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      984 a    683 c    642 g    686 t
ORIGIN
```

```
  Query Match             4.4%;  Score 193;  DB 1;  Length 2995;
  Best Local Similarity   62.0%;  Pred. No. 1.2e-32;
  Matches  323;  Conservative    0;  Mismatches  195;  Indels    3;  Gaps    1;

Qy    1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
           ||||  || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db    1038 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1097

Qy    1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
           ||| |||| ||||  || |||     |  |||||| ||    ||||||| ||| |||||
Db    1098 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 1157

Qy    1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || | |  || | ||  |  ||   |||| |||| ||       |     |||| |
Db    1158 ATTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1217

Qy    1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db    1218 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1277

Qy    1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||||| || ||| || |||| || || || |   |    |||
Db    1278 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1337

Qy    1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
           || || ||     || |||| | ||||||||  | |||||  ||  | | |||
Db    1338 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1397

Qy    1555 aatcagttcaacccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
           |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db    1398 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1457

Qy    1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
           ||  ||     |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db    1458 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1517

Qy    1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
           |||||   |  | | |  |||||  | ||  | || || ||
Db    1518 CACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAA 1558
```

```
RESULT  37
AF235032
LOCUS       AF235032               4659 bp    DNA     linear   BCT 02-MAR-2001
DEFINITION  Neisseria meningitidis IgA1 protease gene, complete cds.
ACCESSION   AF235032
VERSION     AF235032.1  GI:13183004
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4659)
  AUTHORS   Vitovski,S. and Sayers,J.R.
  TITLE     Degenerate specificity of Neisseria meningitidis IgA1 protease
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4659)
  AUTHORS   Vitovski,S. and Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Division MGM, University of Sheffield, F
            floor, Royal Hallamshire Hospital, Sheffield S10 2JF, UK
FEATURES             Location/Qualifiers
     source          1. .4659
                     /organism="Neisseria meningitidis"
                     /strain="NMB"
                     /db_xref="taxon:487"
     CDS             1. .4659
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="AAK15023.1"
                     /db_xref="GI:13183005"
                     /translation="MKTKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHV
                     KGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFV
                     TEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGT
                     PYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDK
                     QEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADKIRQRDNAGTIKGNGEHHWNITFG
                     TNSKIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGIN
                     NDITWLGAGIDVTDGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVI
                     LNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFE
                     HIRNVDEGARIVNHNTSHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRR
                     PIPQGKDLYYKNYRYYALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKN
                     NQRISGFSGFFGEENGKGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLS
                     GRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNV
                     SNITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGN
                     VNLNQNAALVLGKAALWGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNN
                     ASDAQSANQYHTLKINHLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGE
                     PNQEGLNLFDASSVQDRSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRV
                     KPVPSPATNTASQAQTDSAQIAKPQNIVVAPPSPQANQAEEAKRQQAKAEQVKRQQAE
                     AERKSAELAKQKAEAEREARELATRQKAEQERSSAELARRHEKEREAAELSAKQKVEA
                     EREAQALAVRRKAEAEEAKRQAAELARQQEEARKAAELAAKQKAETERKAAEIAEQKA
                     EAEREAAELAKQKAEEEGRQAAQSQPKRRNRRAIPPELSSDATTRALPRIARNSNPDA
                     SDYEKIPLDALEDEDVSESVDTSDKQPQDNTELHEKVETVSLQPRAAQPRAQAAAQAD
                     AVSTNTNSALSDAMASTQSILLDTGASLTRHIAQKSRADAEKNSVWMSNIGYGRDYAS
                     AQYRRFSSKRTQTQIGIDRSLSENMQIGGVLTYSDSQHTFDQASGKNTFVQANLYGKY
```

YLNDAWYVAGDIGAGSLRSRLQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIR
YSRLSSADYKLGNDSVKVSSMSVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKG
GVNVGGNSFAYKADNQQQYSAGAALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF
"

BASE COUNT 1542 a 1074 c 1041 g 1002 t
ORIGIN

Query Match 4.4%; Score 193; DB 1; Length 4659;
Best Local Similarity 62.0%; Pred. No. 1.3e-32;
Matches 323; Conservative 0; Mismatches 195; Indels 3; Gaps 1;

```
Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          ||||  || |||   ||| || |      ||||||| ||||| || || ||  | | |
Db   1084 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1143

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| |||| || |||     |  |||||| ||    ||||||| ||| |||||
Db   1144 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 1203

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || | |  || | ||  |  ||   |||| |||| ||      |    |||| |
Db   1204 ATTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAACGGCGACAGA 1263

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
          || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1264 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1323

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||||| || ||| || |||| || || || |   |    |||
Db   1324 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1383

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||  | |||||  ||  | | |||
Db   1384 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1443

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   1444 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1503

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1504 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1563

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
          |||||   |  | | |  |||||  | ||  | || || ||
Db   1564 CACAACACAAGCCACGCCTCCACGATCACGCTAACGGGTAA 1604
```

RESULT 38
NMA3Z2491/c
LOCUS NMA3Z2491 311321 bp DNA linear BCT 04-DEC-2000
DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome; segment 3/7.
ACCESSION AL162754 AL157959
VERSION AL162754.2 GI:7379424
KEYWORDS .
SOURCE Neisseria meningitidis Z2491.
ORGANISM Neisseria meningitidis Z2491
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 311321)
AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C., Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T., Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K., Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M., Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
TITLE Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
JOURNAL Nature 404 (6777), 502-506 (2000)
MEDLINE 20222556
REFERENCE 2 (bases 1 to 311321)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT Notes:
Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES Location/Qualifiers
source 1. .311321
/organism="Neisseria meningitidis Z2491"
/strain="Z2491"
/db_xref="taxon:122587"
/note="serogroup: A"
RBS 115. .120
gene 129. .587
/gene="NMA0698"
CDS 129. .587
/gene="NMA0698"
/note="NMA0698, possible periplasmic protein, len: 152 aa; unknown, contains a probable N-terminal signal sequence and lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB83985.1"
/db_xref="GI:7379425"
/db_xref="SPTREMBL:Q9JQQ2"
/translation="MLTKLKILLFLFLFVFVLAINLLFFFSSDIESFGNYQFEYVYD
KGWPANYILVMKDGNEGNFDKIISGLVLEYYKEDDNIYFSYIDGQGFASDSCYYKPEI
LYGKIILNKNHIININSMEKNNFLSEDKIMKGTRNWLADPKNKCNIQTLD"
gene 742. .1077
/gene="NMA0699"
CDS 742. .1077
/gene="NMA0699"
/note="NMA0699, pseudogene, probable ABC transporter protein, len: 336 bp; similar to C-termini of many ABC transporters e.g. TR:Q46973 (EMBL:U47048), mtfB, Escherichia coli microcin transport protein (707 aa), fasta scores; E(): 2.6e-19, 53.2% identity in 111 aa overlap. Contains PS00211 ABC transporters family signature. NMA0686 may be the remainder of this pseudogene"
/codon_start=1
/pseudo
/transl_table=11
/product="putative ABC transporter protein (pseudogene)"
misc_feature 868. .912
/gene="NMA0699"
/note="PS00211 ABC transporters family signature"
gene complement(1101. .2327)
/gene="NMA0700"
CDS complement(1101. .2327)
/gene="NMA0700"
/note="NMA0700, possible ribonuclease BN, len: 408 aa; similar to SW:RBN_ECOLI (EMBL:L19201), rbn, Escherichia coli ribonuclease BN (EC 3.1.-.-), fasta scores; E(): 2.3e-28, 34.3% identity in 268 aa overlap. Longer than rbn at the C-terminus. Also similar to TR:O85449 (EMBL:AF067083) Vitreoscilla sp. hypothetical protein (376 aa), fasta scores; E(): 0, 45.0% identity in 333 aa overlap. Contains a region similar to NMA0677, fasta scores; E(): 4.3e-08, 64.4% identity in 45 aa overlap. Contains hydrophobic, possible membrane-spanning regions, as does rbn"
/codon_start=1
/transl_table=11
/product="putative ribonuclease BN"
/protein_id="CAB83987.1"
/db_xref="GI:7379426"
/db_xref="SPTREMBL:Q9JVT9"
/translation="MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLL

```
                ALVPVLTVMVAVASIFPVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTA
                IGSVMLVVTSLMLIRTIDNTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFM
                VGSVQDAALASGAPQWSGALRTAATLTFMTLLLWGLYRFVPNREVPARQAFVGALATA
                FCLETARSLFTWYMGNFDGYRSIYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQ
                GEAFRRGFDSRGRFDDVLKILLLLDAAQKEGKALPVQEFRRHINMGYDELGELLEKLA
                RHGYIYSGRQGWVLKTGADSIELNELFKLFVYRPLPVERDHVNQAVDAVMMPCLQTLN
                MTLAEFDAQAKKQQQS"
misc_feature    1376. .1385
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    complement(2417. .2426)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    2444. .2453
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            2486. .2743
                /gene="NMA0701"
CDS             2486. .2743
                /gene="NMA0701"
                /note="NMA0701, possible pseudogene, len: 258 bp; shows
                weak similarity to part of SW:WRBA_ECOLI (EMBL:M99166),
                wrbA, Escherichia coli Trp repressor binding protein (197
                aa), fasta scores; E(): 0.064, 29.3% identity in 82 aa
                overlap. Also similar to part of TR:O85450 (EMBL:AF067083)
                Vitreoscilla sp. Trp repressor binding protein (fragment)
                (124 aa), fasta scores; E(): 6.6e-14, 56.6% identity in 83
                aa overlap"
                /codon_start=1
                /pseudo
                /transl_table=11
                /product="hypothetical protein NMA0701 (pseudogene)"
misc_feature    complement(2755. .2764)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    2778. .2787
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
RBS             2861. .2864
gene            2871. .3530
                /gene="NMA0702"
CDS             2871. .3530
                /gene="NMA0702"
                /note="NMA0702, len: 219 aa; similar to TR:Q52925
                (EMBL:Z50189), exsB, Rhizobium meliloti putative regulator
                of succinoglycan biosynthesis (not a transcriptional
                regulator) (234 aa), fasta scores; E(): 3.1e-15, 32.7%
                identity in 205 aa overlap. Also similar to many
                hbacterial ypothetical proteins e.g. SW:YBAX_HAEIN
                (EMBL:U32798), HI1191, Haemophilus influenzae hypothetical
                protein (196 aa), fasta scores; E(): 0, 79.2% identity in
                173 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0702"
                /protein_id="CAB83989.1"
                /db_xref="GI:7379427"
                /db_xref="SPTREMBL:Q9JVT8"
                /translation="MSNQKALVIFSGGQDSTTCLIQAIQTYGRENVQAITFQYGQRHA
                VELERARWIAQDLGVKQTVLDLSLMRQITHNALMDDTAAIETAENGVPNTFVDGRNAL
                FLLYAAIYAKGQGIRHIIAGVCETDFSGYPDCRDVFVKSMNVTLNLAMDYDFQIHTPL
                MYLTKAQTWALADEMGALDYIREQTHTCYNGIVGGCRECPSCILRERGLAEYLESKKA
                V"
misc_feature    3522. .3531
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
RBS             3547. .3550
gene            3558. .4076
                /gene="NMA0703"
CDS             3558. .4076
                /gene="NMA0703"
                /note="NMA0703, len: 172 aa; unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0703"
                /protein_id="CAB83990.1"
                /db_xref="GI:7379428"
                /db_xref="SPTREMBL:Q9JVT7"
                /translation="MPKLHMFYLGGNAGRSNIEVHDIQFAVCDDYREAVPALKAAWFG
                DADKIHIDGWQVVEWADGYDIAVSETPKTKMPSENAPRLYFANVGGYRAGQLAEAHAF
                GLFAAATPAEAKQKALQTLLTDSYVQQHKDNLKDVDNLLALDRIGNFHIRLTPNPHGK
                PAEIGFQGYLPI"
misc_feature    3782. .3791
                /gene="NMA0703"
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            4083. .4505
                /gene="NMA0704"
misc_feature    4083. .4502
                /gene="NMA0704"
                /note="Pfam match to entry PF01242 PTPS, 6-pyruvoyl
                tetrahydropterin synthase, score 10.80, E-value 1.7e-06"
CDS             4083. .4505
                /gene="NMA0704"
                /note="NMA0704, len: 140 aa; similar to many hypothetical
                proteins e.g. SW:YB90_HAEIN (EMBL:U32798), HI1190,
                Haemophilus influenzae hypothetical protein (141 aa),
                fasta scores; E(): 5.1e-32, 58.0% identity in 138 aa
                overlap. Shows very weak similarity to eukaryotic
                6-pyruvoyl-tetrahydropterin synthases e.g. SW:PTPS_RAT
                (EMBL:M77850), pts, Rattus norvegicus 6-pyruvoyl
                tetrahydrobiopterin synthase precursor (EC 4.6.1.10) (144
                aa), fasta scores; E(): 0.0062, 27.3% identity in 154 aa
                overlap. Contains Pfam match to entry PF01242 PTPS,
                6-pyruvoyl tetrahydropterin synthase"

 Query Match             4.4%;  Score 193;  DB 1;  Length 311321;
 Best Local Similarity   57.4%;  Pred. No. 2.7e-32;
 Matches  368;  Conservative    0;  Mismatches  270;  Indels    3;  Gaps    1;

Qy    1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
           |||   || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db  188900 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 188841

Qy    1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
           ||| |||| ||||  || |||     || |||||| |     ||||||| ||| || ||
Db  188840 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCTAGGTGCAGGG 188781

Qy    1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |    || | ||  |  ||   |||| |||| ||      |     |||| |
Db  188780 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 188721

Qy    1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||||  |  |  |  | || ||| |  ||||||||
Db  188720 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 188661

Qy    1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || ||||||||||| || ||| || |||| || || || |   |    |||
Db  188660 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 188601

Qy    1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
           || || ||     || ||||| | || |||||  | |||||  ||   |||||
Db  188600 CAGGCTTTCTCCCAAGTCGGCATTGTGAGCGGGCGCGGTACATTGGTGCTGAACAGTCCG 188541

Qy    1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
           | ||| | |||||  | || || |||||||| || || ||||| |||||||||   ||
Db  188540 GACCAGATTAACCCGAATAATCTGTATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 188481

Qy    1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
           ||  ||    |  | ||  | |  || |  ||    |||||||| |||  |||||||||
Db  188480 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 188421

Qy    1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
           |||||  ||||  |  ||||| | ||  | || |||||  |  ||| |  | |
Db  188420 CACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 188361
```

```
Qy   1735 ggcaataacaacagcttggatagcaaaaaagaaattgcctacaacggttggtttggcgag 1794
          |||         |  || | |  ||| ||    |   |  ||| ||  | |||| |
Db 188360 CAAAATCTTTCTGTTTATGAAATCCGAAATGATTACGATGATGACGATTATTATGGCTAT 188301

Qy   1795 aaagatacgaccaaaacgaacgggcggctcaaccttgttta 1835
          |      |  ||| | |    ||    |||   | ||||
Db 188300 TACAGCTATAGAAAACCCATTCCGCAAGGCAAAGATCTTTA 188260


RESULT  39
NMIGAC65
LOCUS       NMIGAC65                1525 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NGC65).
ACCESSION   X82484
VERSION     X82484.1  GI:732871
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1525)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1525)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1525
                     /organism="Neisseria meningitidis"
                     /strain="NGC65"
                     /db_xref="taxon:487"
     gene            1. .1525
                     /gene="iga"
     CDS             <1. .>1525
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57867.1"
                     /db_xref="GI:732872"
                     /db_xref="SPTREMBL:Q51168"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYA
                     VSVKHAKAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWSASNLGRLEDYNM
                     ARFNKFVTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKG
                     YDLRDLSQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALT
                     NYGVLGDSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDN
                     AGTIKGNGEHHWNITFGTNSHIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQ
                     GAGGLFFKGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEIN
                     GTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQVGIVSGRRTLVLNSSNQINPDNLYF
                     GFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITAPQNLSVYEI
                     "
BASE COUNT      510 a    320 c    330 g    365 t
ORIGIN


  Query Match             4.4%;  Score 192.4;  DB 1;  Length 1525;
  Best Local Similarity   60.8%;  Pred. No. 1.5e-32;
  Matches  332;  Conservative    0;  Mismatches  211;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          ||||  || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db    961 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1020

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     || |||||| ||    ||||||| ||| |||||
Db   1021 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGG 1080

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
          || || |    || | ||  |  ||   |||| |||| ||      |      ||| |
Db   1081 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGTGACAGA 1140

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
          || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1141 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1200

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
          | |  || |||||||| ||||| ||| || ||||  | || ||    ||    |||
Db   1201 TTAAAAGTCGGCGACGGCACAGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTC 1260

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||      || ||||| | |||||||  |  ||||  ||  | | |||
Db   1261 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCCGTACATTGGTATTAAATAGTTCA 1320

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   1321 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1380

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||     |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1381 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1440

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaataaagatattgctacaacc 1734
          |||||   || | | |  |||||  | ||  | || |||||    |   ||| |  |  |
Db   1441 CACAACACAGGCCACGCCTCCACAATCACATTAACCGGCAAATCATTAATTACCGCTCCT 1500

Qy   1735 ggcaat 1740
          |||
Db   1501 CAAAAT 1506


RESULT  40
NMIGA139
LOCUS       NMIGA139                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (81139).
ACCESSION   X82477
VERSION     X82477.1  GI:732856
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="81139"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57860.1"
                     /db_xref="GI:732857"
                     /db_xref="SPTREMBL:Q51163"
```

```
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHH
                     WKTTGTNSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYI"
BASE COUNT      498 a    320 c    321 g    350 t
ORIGIN

  Query Match             4.4%;  Score 191.4;  DB 1;  Length 1489;
  Best Local Similarity   61.8%;  Pred. No. 2.5e-32;
  Matches  322;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db    925 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 984

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| |||  || |||     |  |||||| ||    ||||||| ||| |||||
Db    985 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 1044

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
          || || |    || | ||  |  ||   |||| |||| ||      |    |||| |
Db   1045 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1104

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
          || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1105 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1164

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
          | |  || |||||||||| || ||| || |||| || || || |   |    |||
Db   1165 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1224

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||||  | |||||  ||  | | |||
Db   1225 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1284

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   1285 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1344

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1345 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1404

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
          |||||   || | | |  |||||  | ||  | || || ||
Db   1405 CACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAA 1445


RESULT  41
NMIGAC16
LOCUS       NMIGAC16               1525 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NGC16).
ACCESSION   X82485
VERSION     X82485.1  GI:732869
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1525)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1525)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1525
                     /organism="Neisseria meningitidis"
                     /strain="NGC16"
                     /db_xref="taxon:487"
     gene            1. .1525
                     /gene="iga"
     CDS             <1. .>1525
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57868.1"
                     /db_xref="GI:732870"
                     /db_xref="SPTREMBL:Q51167"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYA
                     VSVKHAKAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNM
                     ARFNKFVTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKG
                     YDLRDLSQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALT
                     NYGVLGDSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDN
                     AGTIKGNGEHHWNITFGTNSHIGSTAVRLAGNERDANNGQNVTFEDNGTLVLDQNINQ
                     GAGGLFFKGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEIN
                     GTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYF
                     GFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYI
                     "
BASE COUNT      504 a    322 c    336 g    363 t
ORIGIN

  Query Match             4.4%;  Score 191.4;  DB 1;  Length 1525;
  Best Local Similarity   61.8%;  Pred. No. 2.6e-32;
  Matches  322;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          ||||  || |||   ||| || |       |||||||| ||||| || || ||  | | |
Db    961 GAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 1020

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     || |||||| ||    ||||||| ||| |||||
Db   1021 TTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCTAGGTGCGGGG 1080

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
          || || |    || | ||  |  ||   |||| |||| ||      |    |||| |
Db   1081 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1140

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
          || | ||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1141 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1200

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
          | |  || |||||||||| || ||| || |||| | || ||    ||    |||
Db   1201 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTC 1260

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | |||||||||  | |||||  ||  | | |||
Db   1261 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1320

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| ||  | |||||||| || || ||||| |||||||||   ||
Db   1321 AATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1380

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1381 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1440

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
          |||||   || | | |  |||||  | ||  | || || ||
Db   1441 CACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAA 1481
```

```
RESULT  42
AF012210
LOCUS       AF012210               2995 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4099 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012210
VERSION     AF012210.1  GI:2529503
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Neisseria meningitidis"
                     /strain="Z4099"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup VIII; isolated in China,
                     1975"
     mat_peptide     <1. .>2995
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2995
                     /gene="iga"
                     /allele="iga8"
     CDS             <1. .>2995
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45793.2"
                     /db_xref="GI:12597970"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQHDNAGTIKGYREHHWNITSGTNSKIGSTAVRLAGNE
                     RNANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGNYTVKGANNDITWLGAGIDVADGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     DHTSTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     QKDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFDATQIHGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQERLNLFDASSVRD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRQVKPTHSPATNTASQAQK
                     ATQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      996 a    690 c    632 g    677 t
ORIGIN


  Query Match             4.4%;  Score 191.4;  DB 1;  Length 2995;
  Best Local Similarity   61.8%;  Pred. No. 2.9e-32;
  Matches  322;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    1;

Qy    1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
           ||||  || |||   ||| || |       || ||||| ||||| || || ||  | | |
Db    1038 GAAGACAACGGCACTTTGGTATTGGATCAAAATATCAACCAAGGCGCGGGCGGTCTGTTT 1097

Qy    1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
           ||| ||||  ||| || |||     || |||||| ||    || |||| |||||||||
Db    1098 TTCAAAGGCAATTACACAGTCAAAGGTGCAAATAATGACATCACCTGGCTAGGCGCGGGG 1157

Qy    1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |   || | ||  |  ||   |||| |||| ||      |     |||| |
Db    1158 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1217

Qy    1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db    1218 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1277

Qy    1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || ||||||||||| || ||| || |||| | || ||    ||    |||
Db    1278 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTC 1337

Qy    1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
           || || ||     || ||||| | |||||||  | |||||  ||   || |||
Db    1338 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTGCTAAATAGTTCA 1397

Qy    1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
           |||||| | || ||||| || || |||||||| || || |||| |||||||||   ||
Db    1398 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1457

Qy    1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
           ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db    1458 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1517

Qy    1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
           ||||||  |||| |    |||||  | ||  | || || ||
Db    1518 CACAATACAGACCACACCTCCACAATCACGCTAACGGGTAA 1558


RESULT  43
NMIGA341
LOCUS       NMIGA341               1489 bp    DNA     linear   BCT 12-JUL-1999
DEFINITION  N.meningitidis iga gene (IHN5341).
ACCESSION   X82478
VERSION     X82478.1  GI:732862
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="IHN5341"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /pseudo
                     /transl_table=11
                     /codon_start=1
```

```
                     /product="IgA1 protease"
                     /db_xref="PID:e123968"
BASE COUNT      498 a    319 c    321 g    351 t
ORIGIN


  Query Match             4.3%;  Score 189.8;  DB 1;  Length 1489;
  Best Local Similarity   61.6%;  Pred. No. 5.8e-32;
  Matches  321;  Conservative    0;  Mismatches  197;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |       ||||||||| ||||| || || ||  | | |
Db    925 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 984

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     |  |||||| ||    ||||||| ||| |||||
Db    985 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 1044

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |   || | ||  |  ||   |||| |||| ||      |    |||| |
Db   1045 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1104

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  | | || ||| |  ||||||||
Db   1105 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1164

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||||| || ||| || |||| || || || |   |    |||
Db   1165 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1224

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
           | || ||     || ||||| | |||||||  | |||||  ||  | | |||
Db   1225 TAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1284

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   1285 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1344

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1345 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1404

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
          |||||   || | | |  |||||  | ||  | || || ||
Db   1405 CACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAA 1445


RESULT  44
NMIGAG80
LOCUS       NMIGAG80                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NCG80).
ACCESSION   X82479
VERSION     X82479.1  GI:732883
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="NGC80"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57862.1"
                     /db_xref="GI:732884"
                     /db_xref="SPTREMBL:Q51172"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHH
                     WKTTGTNSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRRTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYI"
BASE COUNT      498 a    321 c    320 g    350 t
ORIGIN


  Query Match             4.3%;  Score 189.8;  DB 1;  Length 1489;
  Best Local Similarity   61.6%;  Pred. No. 5.8e-32;
  Matches  321;  Conservative    0;  Mismatches  197;  Indels    3;  Gaps    1;

Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |       ||||||||| ||||| || || ||  | | |
Db    925 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 984

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     |  |||||| ||    ||||||| ||| |||||
Db    985 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 1044

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |   || | ||  |  ||   |||| |||| ||      |    |||| |
Db   1045 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1104

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||  |  |  | | || ||| |  ||||||||
Db   1105 TTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1164

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||||| || ||| || |||| || || || |   |    |||
Db   1165 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGACAAAAAAGTC 1224

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
           || || ||     || ||||| | |||||||  |  |||| ||  | | |||
Db   1225 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCCGTACATTGGTATTAAATAGTTCA 1284

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   1285 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1344

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
          ||  ||    |  | ||  | |  || |  ||    |||||||| |||   |||||||||
Db   1345 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1404

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
          |||||   || | | |  |||||  | ||  | || || ||
Db   1405 CACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAA 1445


RESULT  45
NMIGAF48
LOCUS       NMIGAF48                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (HF48).
ACCESSION   X82480
VERSION     X82480.1  GI:732875
```

```
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="HF48"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57863.1"
                     /db_xref="GI:732876"
                     /db_xref="SPTREMBL:Q57035"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQKADSNQKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHTSTITLTGKSLITNPNSLSVHSI"
BASE COUNT      496 a    319 c    325 g    349 t
ORIGIN
```

Query Match 4.3%; Score 188.2; DB 1; Length 1489;
Best Local Similarity 61.4%; Pred. No. 1.3e-31;
Matches 320; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

```
Qy   1198 gaaggaaaaggcgaattgatacttaccagcaacatcaatcaaggtgctggaggattatat 1257
          |||   || |||   ||| || |       ||||||||| ||||| || || ||  |·| |
Db    925 GAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGGCGGTCTGTTT 984

Qy   1258 ttccaaggagattttacggtctcgcctgaaaataacgaa---acttggcaaggcgcgggc 1314
          ||| |||| ||||  || |||     |  |||||| ||    ||||||| ||| |||||
Db    985 TTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCTAGGTGCGGGG 1044

Qy   1315 gttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgc 1374
           || || |    || | ||  |  ||   |||| |||| ||       |     |||| |
Db   1045 ATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAATGGCGACAGA 1104

Qy   1375 ctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcg 1434
           || | ||||||||||||||||||  |  |  |  | | || ||| |  ||||||||
Db   1105 TTGGCGAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAACCAAGGGCAA 1164

Qy   1435 atcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaa 1494
           | |  || |||||||||||| || ||| || ||||  | || ||    ||    |||
Db   1165 TTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAACCAAAAAGTC 1224

Qy   1495 caagcctttagtgaaatcggcttggtcagcggcaggggtacggtgcaactgaatgccgat 1554
          || || ||     || ||||| | ||||||||  | |||||  ||  | | |||
Db   1225 CAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATTAAATAGTTCA 1284

Qy   1555 aatcagttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaac 1614
          |||||| | || ||||| || || |||||||| || || ||||| |||||||||   ||
Db   1285 AATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTTGGATGCCAAT 1344

Qy   1615 gggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgtcaac 1674
           ||  ||    |  | ||  | |  || |  ||    |||||||| |||  |||||||||
Db   1345 GGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCGCATTGTCAAC 1404

Qy   1675 cacaatcaagacaaagaatccaccgttaccattacaggcaa 1715
          |||||   || | |    |||||  | ||  | || || ||
Db   1405 CACAACACAGGCCACACCTCCACAATCACGCTAACGGGTAA 1445
```

Search completed: June 30, 2002, 11:42:53
Job time: 66391 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 11:42:53 ; Search time 27321.5 Seconds
(without alignments)
3331.827 Million cell updates/sec

Title: US-09-303-518D-651
Perfect score: 4350
Sequence: 1 atgaaaacaaccgacaaacg.....aattaggctaccgctggtaa 4350

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4288	98.6	4350	6	A96618	A96618 Sequence 65
	2	4288	98.6	349061	1	NMA2Z2491	AL162753 Neisseria
	3	4284.8	98.5	4350	1	NME296276	AJ296276 Neisseria
	4	3508	80.6	4374	6	A96616	A96616 Sequence 64
	5	3508	80.6	4374	6	AX044007	AX044007 Sequence
c	6	3508	80.6	10074	1	AE002547	AE002547 Neisseria
c	7	3508	80.6	349980	6	AX044034	AX044034 Sequence
	8	3500	80.5	4374	1	NME242535	AJ242535 Neisseria
	9	3490.2	80.2	4386	1	NME296277	AJ296277 Neisseria
	10	3385.8	77.8	4407	6	A96620	A96620 Sequence 65
	11	2135.4	49.1	4355	6	A96614	A96614 Sequence 64
	12	1359.4	31.3	4317	1	HIU11024	U11024 Haemophilus
	13	1348	31.0	4319	6	AR157320	AR157320 Sequence
c	14	1255.8	28.9	13818	1	U32710	U32710 Haemophilus
	15	550	12.6	578	1	NLIGA3327	AJ001740 Niesseria
	16	540	12.4	584	1	NLIGA3354	AJ001739 Niesseria
	17	517.6	11.9	572	1	NLIGA823	AJ001737 Niesseria
	18	332.8	7.7	572	1	NLIGA3293	AJ001738 Niesseria
	19	197.6	4.5	1492	1	NMIGA476	X82481 N.meningiti
	20	197.6	4.5	2983	1	AF012208	AF012208 Neisseria
	21	197.6	4.5	2995	1	AF012209	AF012209 Neisseria
	22	197.6	4.5	4659	1	AF235032	AF235032 Neisseria
	23	196	4.5	2995	1	AF012210	AF012210 Neisseria
	24	192.8	4.4	1489	1	NMIGA771	X82468 N.meningiti
	25	192.8	4.4	1489	1	NMIGAF48	X82480 N.meningiti
	26	192.8	4.4	1489	1	NMIGAF96	X82475 N.meningiti
	27	192.8	4.4	1489	1	NMIGATH2	X82469 N.meningiti
	28	192.8	4.4	1525	1	NMIGAG117	X82483 N.meningiti
	29	192.8	4.4	2980	1	AF012211	AF012211 Neisseria
	30	192.8	4.4	2992	1	AF012207	AF012207 Neisseria
	31	192.8	4.4	5261	1	NMIGAF13	X82474 N.meningiti
c	32	191.4	4.4	10485	1	AE002424	AE002424 Neisseria
c	33	191.4	4.4	349980	6	AX044030	AX044030 Sequence
	34	191.2	4.4	1489	1	NMIGA139	X82477 N.meningiti
	35	191.2	4.4	1489	1	NMIGA894	X82476 N.meningiti
	36	191.2	4.4	1525	1	NMIGAC16	X82485 N.meningiti
	37	191.2	4.4	2992	1	AF012206	AF012206 Neisseria
	38	191.2	4.4	4899	1	NGIGAPRO	X04835 Neisseria g
	39	191.2	4.4	4899	6	A02796	A02796 N.gonorrhoe
	40	191.2	4.4	4899	6	A12416	A12416 IgA-proteas
	41	191.2	4.4	4899	6	AR079202	AR079202 Sequence
	42	191.2	4.4	8393	1	NGCOMLGN	Z49895 N.gonorrhoe
	43	189.6	4.4	1489	1	NMIGA341	X82478 N.meningiti
	44	189.6	4.4	1489	1	NMIGAG80	X82479 N.meningiti
	45	189.6	4.4	1525	1	NMIGAC65	X82484 N.meningiti
	46	188	4.3	1522	1	NMIGAO93	X82482 N.meningiti
	47	185.8	4.3	6021	1	HEAIGA1PTC	M87491 Haemophilus
	48	183.2	4.2	1489	1	NMIGA027	X82472 N.meningiti
	49	183.2	4.2	1489	1	NMIGA117	X82470 N.meningiti
	50	183.2	4.2	1489	1	NMIGA159	X82471 N.meningiti
	51	183.2	4.2	1525	1	NMIGA166	X82486 N.meningiti
	52	183.2	4.2	2980	1	AF012203	AF012203 Neisseria
	53	183.2	4.2	2980	1	AF012204	AF012204 Neisseria
	54	183.2	4.2	2980	1	AF012205	AF012205 Neisseria
c	55	183.2	4.2	311321	1	NMA3Z2491	AL162754 Neisseria
	56	181.6	4.2	1489	1	NMIGAF54	X82473 N.meningiti
	57	180.2	4.1	5536	1	HEAIGA1PTA	M87489 Haemophilus
	58	180.2	4.1	5657	1	HIIGA1P	X59800 Haemophilus
c	59	180.2	4.1	15212	1	U32779	U32779 Haemophilus
	60	177.4	4.1	11288	1	AE002549	AE002549 Neisseria
	61	177.4	4.1	172325	6	AX044035	AX044035 Sequence
	62	177.4	4.1	349980	6	AX044034	AX044034 Sequence
	63	168.2	3.9	5088	1	HEAIGA1PTB	M87490 Haemophilus
	64	161.6	3.7	5719	1	AF369907	AF369907 Haemophil
	65	158	3.6	5096	1	HEAIGA1PTD	M87492 Haemophilus
	66	158	3.6	5096	1	HIGA	X64357 H.influenza
c	67	144	3.3	213494	1	SFPWR100	AL391753 Shigella
	68	142.4	3.3	4484	1	SFSEPAGN	Z48219 Shigella fl
	69	138.2	3.2	4113	6	AX276281	AX276281 Sequence
c	70	138.2	3.2	48254	6	AX276220	AX276220 Sequence
c	71	137.2	3.2	37710	1	ECO278144	AJ278144 Escherich
	72	135	3.1	4394	1	SFU35656	U35656 Shigella fl
	73	135	3.1	5639	1	AF097644	AF097644 Escherich

c	74	135	3.1	51334	1	AF200692	AF200692 Shigella
	75	132.6	3.0	1390	1	HIIGA635	X82488 H.influenza
	76	122.6	2.8	9689	7	AF151674	AF151674 Bacteriop
	77	122.6	2.8	11279	7	AF151091	AF151091 Prophage
	78	122	2.8	4128	6	AX275950	AX275950 Sequence
	79	122	2.8	5208	1	AF422145	AF422145 Escherich
c	80	122	2.8	7654	6	AX275945	AX275945 Sequence
	81	113.8	2.6	1413	1	HIIGA284	X82487 H.influenza
	82	109.4	2.5	10587	1	AF218073	AF218073 Escherich
	83	109.4	2.5	12606	1	ECAJ3631	AJ223631 Escherich
	84	97.4	2.2	1380	1	HIDNAIGA	X86103 H.influenza
	85	96	2.2	1477	1	HIIGAA66	X82467 H.influenza
c	86	75.2	1.7	1098	1	SF2AETX1	Z47381 S.flexneri
c	87	61.4	1.4	7218	6	I66494	I66494 Sequence 14
	88	60.4	1.4	2070	5	BSU43200	U43200 Boreogadus
	89	58.4	1.3	144	6	A12818	A12818 IgA-proteas
c	90	55.4	1.3	125020	9	AF429315	AF429315 Homo sapi
c	91	55	1.3	18471	1	AE007047	AE007047 Mycobacte
c	92	55	1.3	36021	1	MTCY359	Z83859 Mycobacteri
	93	54.2	1.2	125020	9	AF429315	AF429315 Homo sapi
c	94	52.6	1.2	15776	1	AE007069	AE007069 Mycobacte
c	95	52.6	1.2	37586	1	MTCY270	Z95388 Mycobacteri
c	96	52.6	1.2	37586	6	AX191745	AX191745 Sequence
	97	52.2	1.2	10732	6	E32986	E32986 Gene encodi
	98	52	1.2	1316	14	AY034637	AY034637 Equine he
	99	51	1.2	313450	1	AL596170	AL596170 Listeria
	100	50.4	1.2	1226	14	AY034642	AY034642 Equine he

ALIGNMENTS

```
RESULT   1
A96618
LOCUS       A96618                   4350 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 651 from Patent WO9924578.
ACCESSION   A96618
VERSION     A96618.1  GI:6780201
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4350)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 651 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4350
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     1225 a   1161 c   1060 g    842 t     62 others
ORIGIN
```

Query Match 98.6%; Score 4288; DB 6; Length 4350;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTNTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy    361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AACGTTGATTTTGGTGCGGAAGGAAGNAATCCCGATCAGCACCGTTTTTCTTACCAAATT 420

Qy    421 gtgaaaagaaataattataagcctgacaattcacacccttacaacggcgattancatatg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACAACGGCGATTANCATATG 480

Qy    481 ccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgagg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATGAGG 540

Qy    541 gggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacac 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACAC 600

Qy    601 cactattggcgttatgatgatgacaaacacggcgatttatcctactccggcgcatggtta 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CACTATTGGCGTTATGATGATGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTA 660

Qy    661 attggcggcaatacacatatgcagggttggggaaataatggcgtanttagtttgagcggc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAATGGCGTANTTAGTTTGAGCGGC 720

Qy    721 gatgtgcgccatgccaacgactatggccctatgccgattgcaggtgcggcaggcgacagc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAGC 780

Qy    781 ggttcgccaatgtttatttatgacaaaacaaacaataaatggctgctcaacggagtttta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCTCAACGGAGTTTTA 840

Qy    841 caaaccggctacccttattccggcagggaaaacggtttccagctgatacgcaaagattgg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTCCAGCTGATACGCAAAGATTGG 900

Qy    901 ttctacgatgacatttacagaggcgatacacataccgtctntttttgaaccgcgcagtaac 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTCTACGATGACATTTACAGAGGCGATACACATACCGTCTNTTTTGAACCGCGCAGTAAC 960

Qy    961 ggacatttttcctttacatccaacaacaacggtacgggtacggtaacagaaaccaacgaa 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GGACATTTTTCCTTTACATCCAACAACAACGGTACGGGTACGGTAACAGAAACCAACGAA 1020

Qy   1021 aaggtntccaatccaaagcttaaagtacagacagtccgactgtttgacgaatctttgaat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAAT 1080

Qy   1081 gaaactgataaagaaccagtttacgcggcaggggtgttaatcagtaccgtccaaggtta 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAAACTGATAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCCAAGGTTA 1140

Qy   1141 aacaacggtgaaaacctttctttttatcgattacggcaacggcaaactcatcttatcaaac 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAAC 1200

Qy   1201 aacatcaaccaaggcgcgggcggtttgtattttgaaggtgattttacggtctcgcctgaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCTGAA 1260

Qy   1261 aacaacgaaacgtggcaaggcgcgggcgttcatatcagtgaagacagtaccgttacttgg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1261 AACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGG 1320

Qy   1321 aaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgtt 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTT 1380

Qy   1381 caagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtacagtcattttggat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGAT 1440

Qy   1441 cagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcggcttgntcagcggc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGNTCAGCGGC 1500

Qy   1501 aggggtacggtgcaactgaatgccgataatcagttcaaccccgacaaaactctatttcggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGC 1560

Qy   1561 tttcgcggcggacgtttggatttaaacgggcattcgctttcgttccaccgtattcaaaat 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAAT 1620

Qy   1621 accgatgaaggggcgatgattgncnatcataatgccacaacaacatccaccgttaccatt 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ACCGATGAAGGGGCGATGATTGNCNATCATAATGCCACAACAACATCCACCGTTACCATT 1680

Qy   1681 acagggaatgaaagtattacacaaccgagtggtaagaatatcaatagacttaattacagc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ACAGGGAATGAAAGTATTACACAACCGAGTGGTAAGAATATCAATAGACTTAATTACAGC 1740

Qy   1741 aaagaaattgcctacaacggttggtttggcgagaaagatacgaccaaaacgaacgggcgg 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGG 1800

Qy   1801 ctcaaccttgtttaccagcccgccgcagaagaccgcacccngctgctttccggcggaaca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCNGCTGCTTTCCGGCGGAACA 1860

Qy   1861 aatttaaacggcaacatcacgcaaacaaacggcaaactgtttttcagcggcagaccgaca 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACA 1920

Qy   1921 ccgcacgcctacaatcatttaggaagcgggtggtcaaaaatggaaggtatcccacaagga 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 1980

Qy   1981 gaaatcgtgtgggacaacgactggatcnaccgcacgtttaaagcggaaaatttccatatt 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GAAATCGTGTGGGACAACGACTGGATCNACCGCACGTTTAAAGCGGAAAATTTCCATATT 2040

Qy   2041 cagggcgggcaggcggtgatttcccgcaatgttgccaaagtggaaggcgattgncatttg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGNCATTTG 2100

Qy   2101 agcaatcacgcccaagcagtttttggtgtcgcaccgcatcaaagccatacaatctgtaca 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATCTGTACA 2160

Qy   2161 cgttcggactggacnggtctgacaaattgtgtcgaanaaancattaccgacgataaagtg 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CGTTCGGACTGGACNGGTCTGACAAATTGTGTCGAANAAANCATTACCGACGATAAAGTG 2220

Qy   2221 attgcttcattgactaagacngacntnagcggcantgtnagnctnnccnatnacgntnnt 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATTGCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGNCTNNCCNATNACGNTNNT 2280

Qy   2281 tnaaanctcncngggcntgcnncactnaanggcaatcttagtgcaaatggcgatacacgt 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TNAAANCTCNCNGGGCNTGCNNCACTNAANGGCAATCTTAGTGCAAATGGCGATACACGT 2340

Qy   2341 tatacagtcagccacaacgccacccaaaacggcaaccttagcctcgtgggcaatgcccaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAA 2400

Qy   2401 gcaacatttaatcaagccacattaaacggcaacncatcggnttcgggcaatgcttcattt 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAACATTTAATCAAGCCACATTAAACGGCAACNCATCGGNTTCGGGCAATGCTTCATTT 2460

Qy   2461 aatctaagcaacaacgccgcacaaaacggcagtctgacgctttccgacaacgctaaggca 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 AATCTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCA 2520

Qy   2521 aacgtaagccattccgcactcaacggcaatgtctccctagccgataaggcagtattccat 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 AACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT 2580

Qy   2581 tttgaaaacagccgctttaccggacaactcagcggcagcaaggganacagcattacactta 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TTTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGANACAGCATTACACTTA 2640

Qy   2641 aaagacagcgaatggacgctgccgtcaggcacggaattaggcaatttaaaccttgacaac 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAAC 2700

Qy   2701 gccaccattacactcaattccgcctatcgccacgatgctgcaggcgcgcaaaccggcagn 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GCCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGN 2760

Qy   2761 gtgtcagacacgccgcgccgccgttcgcgccgttccctattatccgttacaccgccaact 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACT 2820

Qy   2821 tcggtagaatcccgtttcaacacgctgacggtaaacggcaaattgaacngtcaaggaaca 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 TCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACNGTCAAGGAACA 2880

Qy   2881 ttccgctttatgtcggaactcttcggctaccgaagcgacaaattgaagctggcggaaagt 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGT 2940

Qy   2941 tccgaaggnacttacaccttggcggtcaacaataccggcaacgaacccgtaagcctcgat 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TCCGAAGGNACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGAT 3000

Qy   3001 caattgacggtagtggaagggaaagacaacaaaccgctgtccgaaaaccttaatttcacc 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 CAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACC 3060

Qy   3061 ctgcaaaacgaacacgtcgatgccggcgcgtggcgttaccaactcatccgcaaagacggc 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 CTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3120

Qy   3121 gagttccgcctgcataatccggtcaaagaacaagagctttccgacaaactcggcaaggca 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCA 3180

Qy   3181 gaagccaaaaaacaggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcggcc 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCC 3240

Qy   3241 gggcgcgatgccgccgaaaagacagaaagcgttgccgaaccggcccggcnggcaggcggg 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCNGGCAGGCGGG 3300

Qy   3301 gaaaatgtcggcattatgcaggcggaggaagagagaaaaaacgggtgcaggcggataaagac 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC 3360

Qy   3361 agcgcnttggcgaaacagcgcgaagcggaaacccggccggntaccaccgccttcccccgc 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 AGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGNTACCACCGCCTTCCCCCGC 3420

Qy   3421 gcccgcngcgcccgccgggatttgccgcaaccgcagccccaaccgcaacctcaaccccaa 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GCCCGCNGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAA 3480
```

```
Qy   3481 ccgcagcgcgacctgatnagccgttatgccaatagcggtttgagtgaattttccgccacg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CCGCAGCGCGACCTGATNAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACG 3540

Qy   3541 ctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagaccgccgc 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGC 3600

Qy   3601 aacgcngtttggacaagcngcatccggnacaccaaacactaccgttcgcaagatttccgc 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 AACGCNGTTTGGACAAGCNGCATCCGGNACACCAAACACTACCGTTCGCAAGATTTCCGC 3660

Qy   3661 gcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggcagcgggg 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG 3720

Qy   3721 cgcgtcggcatcctgttttcgcacaaccggaccgaaaacancttcgacgacggcatcggc 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 CGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACANCTTCGACGACGGCATCGGC 3780

Qy   3781 aactcggcacggcttgcccacggcgccgttttcgggcaatacggcatcggcaggttcgac 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGAC 3840

Qy   3841 atcggcatcagcacgggcgcgggttttagcagcggcantctntcagacggcatcggaggc 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 ATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCANTCTNTCAGACGGCATCGGAGGC 3900

Qy   3901 aaaatccgccgccgcgtgctgcattacggcattcaggcacgataccgcgccggtttcggc 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 AAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGC 3960

Qy   3961 ggattcggcatcgaaccgtacatcggcgcaacgcgctatttcgtccaaaaagcggattac 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTAC 4020

Qy   4021 cgctacgaaaacgtcaatatcgccacccccggtcttgcgttcaaccgntaccgngcgggc 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 CGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGNTACCGNGCGGGC 4080

Qy   4081 attaaggcagattattcattcaaaccggcgcaacacatntccatcacnccttatttnagc 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 ATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATNTCCATCACNCCTTATTTNAGC 4140

Qy   4141 ctgtcctataccgatgccgcttcgggcaaagtccgaacacgcgtcaataccgcngtattg 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 CTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCNGTATTG 4200

Qy   4201 gctcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatcaaaggt 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 GCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGT 4260

Qy   4261 ttcacgctgtccntccacgctgccgccgccaaaggnccgcaactggaagcgcaacacagc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 TTCACGCTGTCCNTCCACGCTGCCGCCGCCAAAGGNCCGCAACTGGAAGCGCAACACAGC 4320

Qy   4321 gcgggcatcaaattaggctaccgctggtaa 4350
          |||||||||||||||||||||||||||||
Db   4321 GCGGGCATCAAATTAGGCTACCGCTGGTAA 4350
```

```
RESULT   2
NMA2Z2491
LOCUS       NMA2Z2491              349061 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 2/7.
ACCESSION   AL162753 AL157959
VERSION     AL162753.2  GI:7379120
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349061)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 349061)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .349061
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     repeat_unit     complement(24. .206)
                     /note="ATR repeat; hmmfs hit to HMM ATR (1 - 183), score:
                     310.39"
                     /label=ATR
     RBS             209. .212
     gene            220. .681
                     /gene="NMA0368"
     CDS             220. .681
                     /gene="NMA0368"
                     /note="NMA0368, probable integral membrane protein, len:
                     153 aa; contains four probable transmembrane domains"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAB83669.1"
                     /db_xref="GI:7379121"
                     /db_xref="SPTREMBL:Q9JWH7"
                     /translation="MQEQNRKSSFPIVMLLVSVALWIASLSNVAFYLGNHGSMEGLTV
                     LILGSIFASLDIRYCAVYANYVWLAAIVLLALRKKVVPVHAAFWGLALVAFSVKAVYV
                     DEAGNTSDIVRYGAGFYLWYAAFAVASIGTFAGKNKERKAASAADGTKNDV"
     gene            671. .1492
                     /gene="hemK"
     CDS             671. .1492
                     /gene="hemK"
                     /note="NMA0369, hemK, HemK protein, len: 273 aa; similar
                     to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
                     scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
                     PS00092 N-6 Adenine-specific DNA methylases signature"
                     /codon_start=1
                     /transl_table=11
                     /product="HemK protein"
                     /protein_id="CAB83670.1"
                     /db_xref="GI:7379122"
                     /db_xref="SPTREMBL:Q9JWH6"
                     /translation="MTFDKWLGLSKLPKNEARMLLQYVSEYTRVQLLTRGGEEMPDEV
                     RQRADRLAQRRLNGEPVAYILGAREFYGRRFTVNPSVLIPRPETEHLVEAVLARLPEN
                     GRVWDLGTGSGAVAVTVALERPDAFVRASDISPPALETARKNAADLGARVEFAYGSWF
                     DTDMPSEGKWDIIVSNPPYIENGDKHLSQGDLRFEPQIALTDFSDGLSCIRTLAQGAP
                     DRLAEGGFLLLEHGFDQGAAVRGVLAENGFSGVETLPDLAGLDRVTLGKYMKHLK"
     misc_feature    831. .840
                     /gene="hemK"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    1162. .1171
                     /gene="hemK"
                     /note="Core DNA uptake sequence: gccgtctgaa"
```

```
                /label=DUS
misc_feature    1187. .1207
                /gene="hemK"
                /note="PS00092 N-6 Adenine-specific DNA methylases
                signature"
misc_feature    complement(1279. .1288)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
RBS             1565. .1568
gene            1578. .2966
                /gene="NMA0370"
CDS             1578. .2966
                /gene="NMA0370"
                /note="NMA0370, probable integral membrane protein, len:
                462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
                P44640 hypothetical protein HI0325 (450 aa), fasta scores;
                E(): 0, 49.8% identity in 464 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="putative integral membrane protein"
                /protein_id="CAB83671.1"
                /db_xref="GI:7379123"
                /db_xref="SPTREMBL:Q9JWH5"
                /translation="MNAVVVAVIVMLVLSLSRVHVVLSLTIGAFVGGAVAGMPLQNIA
                DAAGQVSQAGIIPVFNKGLEGGAKIALSYAMLGAFAMAITHSGLPQQLAGAVVRKLNR
                GGMPDSVRSGEGVVKWLLLSIILVMGMMSQNIIPIHIAFIPMIVPPLLLVFNRLKIDR
                RLIACVITFGLVTTYMFLPYGFGAIFLNEILLGNIHSAAPQLDVKNINVMAAMAIPAL
                GMLAGLLLAFVHYRKPRLYQSNNADTAGNADAANRPQPSAYRSLAAAVAIAVCFAIQL
                MYEDSLVLGAMLGFAVFMMLGVINRDKANDVFGEGIKMMAMVGFIMIAAQGFAAVMNA
                TGHIQPLVESSMAIFGNSKGMAALAMLVVGLLVTMGIGSSFSTLPIIAAIYVPLCVGL
                GFSPLATVAIVGTAGALGDAGSPASDSTLGPTMGLNADGQHDHIRDSVIPTFIHYNIP
                LLIAGWIAAMVL"
gene            2967. .3191
                /gene="slyX"
CDS             2967. .3191
                /gene="slyX"
                /note="NMA0371, slyX, SLYX protein homolog, len: 74 aa;
                simlar to SLYX_HAEIN P44759 SLYX protein homolog (73 aa),
                fasta scores; E(): 0.0017, 33.8% identity in 74 aa
                overlap, and SLYX_ECOLI P30857 SLYX protein (72 aa), fasta
                scores; E(): 0.24, 32.4% identity in 68 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="SLYX protein homolog"
                /protein_id="CAB83672.1"
                /db_xref="GI:7379124"
                /db_xref="SPTREMBL:Q9JWH4"
                /translation="MDAVQEFEHRITELEIQSALQEDVIAGLNAMVAELRQTLDLQQA
                QLRLLYQKMQDRNPDAQEPYSLRDEIPPHY"
gene            complement(3271. .3423)
                /gene="NMA0372"
CDS             complement(3271. .3423)
                /gene="NMA0372"
                /note="NMA0372, unknown, questionable CDS, len: 50 aa"
                /codon_start=1
                /transl_table=11
                /product="very hypothetical protein NMA0372"
                /protein_id="CAB83673.1"
                /db_xref="GI:7379125"
                /db_xref="SPTREMBL:Q9JWH3"
                /translation="MFGRLSPLGSDGFFLAVYEYPVGFICLQGKTAYFRSGGKRFHRI
                KGILSD"
misc_feature    3562. .3571
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(3586. .4356)
                /gene="thiF"
CDS             complement(3586. .4356)
                /gene="thiF"
                /note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
                similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
                fasta scores; E(): 0, 43.1% identity in 246 aa overlap,
                and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
                protein. (249 aa), fasta scores; E(): 0, 43.9% identity in
                244 aa overlap (note that N.m. does not have orthologs of
                any other molybdopterin biosynthesis proteins). Contains
                Pfam match to entry PF00899 ThiF_family, ThiF family"
                /codon_start=1
                /transl_table=11
                /product="ThiF protein"
                /protein_id="CAB83674.1"
                /db_xref="GI:7379126"
                /db_xref="SPTREMBL:Q9JWH2"
                /translation="MTTTEHDNDDAFLLRYSRHILLDEIGIEGQQKLSAAHILVVGCG
                GLGAAALPYLAASGIGTLTIADSDTVELHNLQRQVAFDEGDVGKLKTEALADRLRHIN
                HTVDVRTINEKLDGCRLTGLVQAADIVLDCCDNYATRQAVNRACVQAKTPLVSGAAVR
                FEGQLAVYRPDLPDSPCYACLFDGGSASDGICSLFGVFSPLVGIIGSTQAAEALKILL
                DAGEPSHGRLAVYRALEGGWQYFDLPRNPECPVCGAER"
misc_feature    3787. .3796
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    complement(3847. .4257)
                /gene="thiF"
                /note="Pfam match to entry PF00899 ThiF_family, ThiF
                family, score 186.60, E-value 4.1e-52"
misc_feature    complement(4418. .4427)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            4471. .7224
                /gene="ppc"
CDS             4471. .7224
                /gene="ppc"
                /EC_number="4.1.1.31"
                /note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
                917 aa; simlar to many e.g. CAPP_RHOPA O32483
                phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
                fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
                Contains 2x Pfam match to entry PF00311 PEPcase,
                Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
                site motif A (P-loop), and PS00393 Phosphoenolpyruvate
                carboxylase active site 2"
                /codon_start=1
                /transl_table=11
                /product="phosphoenolpyruvate carboxylase"
                /protein_id="CAB83675.1"
                /db_xref="GI:7379127"
                /db_xref="SPTREMBL:Q9JWH1"
```

```
Query Match             98.6%;  Score 4288;  DB 1;  Length 349061;
Best Local Similarity   98.6%;  Pred. No. 0;
Matches 4288;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94066 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 94125

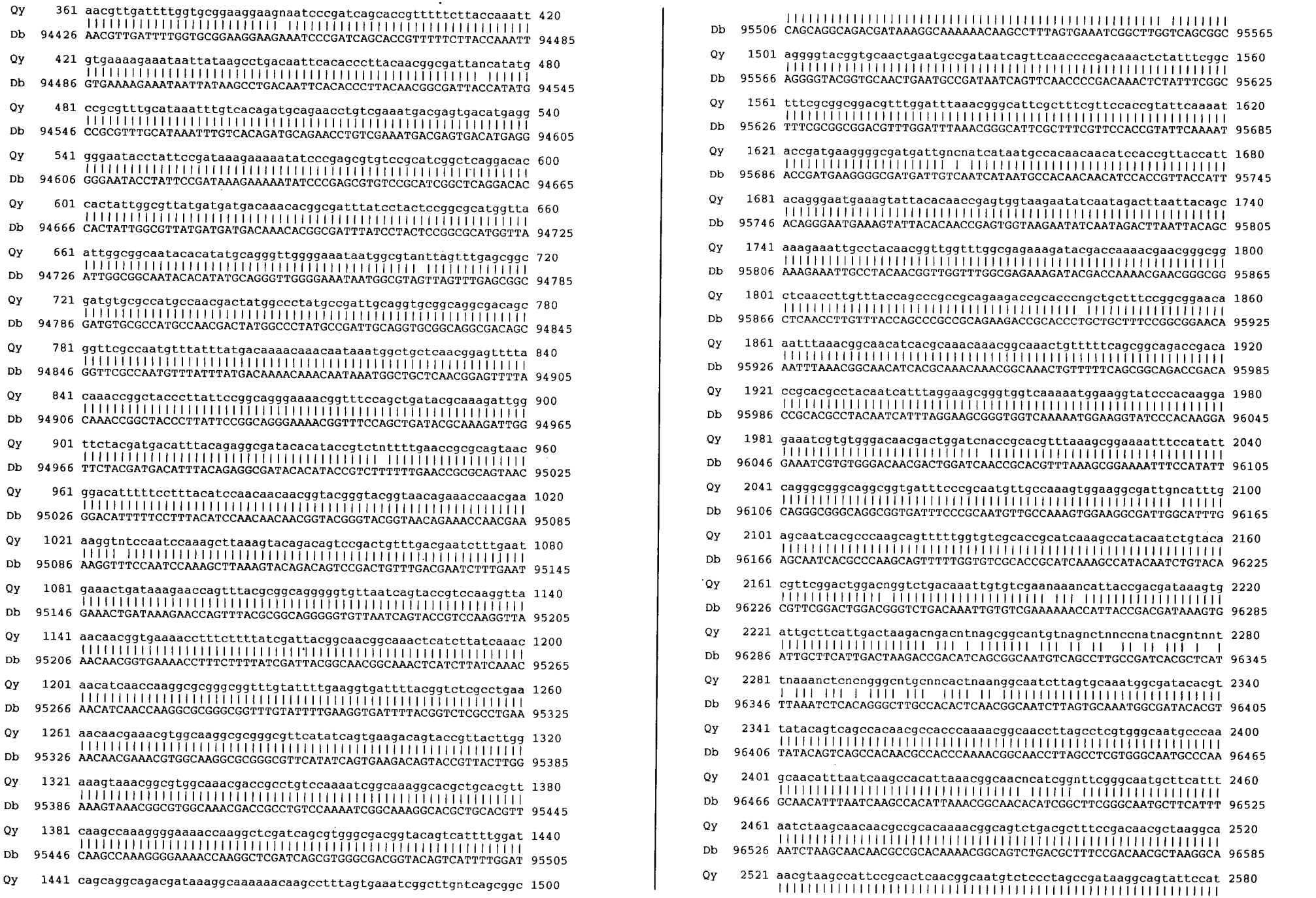
Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94126 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 94185

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94186 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 94245

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  94246 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 94305

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94306 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 94365

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  94366 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 94425
```


Qy 361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
Db 94426 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAGCACCGTTTTTCTTACCAAATT 94485
Qy 421 gtgaaaagaaataattataagcctgacaattcacacccttacaacggcgattancatatg 480
Db 94486 GTGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACAACGGCGATTACCATATG 94545
Qy 481 ccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgagg 540
Db 94546 CCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATGAGG 94605
Qy 541 gggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacac 600
Db 94606 GGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACAC 94665
Qy 601 cactattggcgttatgatgatgacaaacacggcgatttatcctactccggcgcatggtta 660
Db 94666 CACTATTGGCGTTATGATGATGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTA 94725
Qy 661 attggcggcaatacacatatgcagggttggggaaataatggcgtanttagtttgagcggc 720
Db 94726 ATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAATGGCGTAGTTAGTTTGAGCGGC 94785
Qy 721 gatgtgcgccatgccaacgactatggccctatgccgattgcaggtgcggcaggcgacagc 780
Db 94786 GATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAGC 94845
Qy 781 ggttcgccaatgtttatttatgacaaaacaaacaataaatggctgctcaacggagtttta 840
Db 94846 GGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCTCAACGGAGTTTTA 94905
Qy 841 caaaccggctacccttattccggcagggaaaacggtttccagctgatacgcaaagattgg 900
Db 94906 CAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTCCAGCTGATACGCAAAGATTGG 94965
Qy 901 ttctacgatgacatttacagaggcgatacacataccgtctnttttgaaccgcgcagtaac 960
Db 94966 TTCTACGATGACATTTACAGAGGCGATACACATACCGTCTTTTTTGAACCGCGCAGTAAC 95025
Qy 961 ggacatttttccttacatccaacaacaacggtacgggtacggtaacagaaaccaacgaa 1020
Db 95026 GGACATTTTTCCTTTACATCCAACAACAACGGTACGGGTACGGTAACAGAAACCAACGAA 95085
Qy 1021 aaggtntccaatccaaagcttaaagtacagacagtccgactgtttgacgaatctttgaat 1080
Db 95086 AAGGTTTCCAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAAT 95145
Qy 1081 gaaactgataaagaaccagtttacgcggcaggggtgttaatcagtaccgtccaaggtta 1140
Db 95146 GAAACTGATAAAGAACCAGTTTACGCGGCAGGGGTGTTAATCAGTACCGTCCAAGGTTA 95205
Qy 1141 aacaacggtgaaaacctttcttttatcgattacggcaacggcaaactcatcttatcaaac 1200
Db 95206 AACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAAC 95265
Qy 1201 aacatcaaccaaggcgcgggcggtttgtattttgaaggtgattttacggtctcgcctgaa 1260
Db 95266 AACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCTGAA 95325
Qy 1261 aacaacgaaacgtggcaaggcgcgggcgttcatatcagtgaagacagtaccgttacttgg 1320
Db 95326 AACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGG 95385
Qy 1321 aaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgtt 1380
Db 95386 AAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTT 95445
Qy 1381 caagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtacagtcattttggat 1440
Db 95446 CAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGAT 95505
Qy 1441 cagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcggcttgntcagcggc 1500
Db 95506 CAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGC 95565
Qy 1501 aggggtacggtgcaactgaatgccgataatcagttcaaccccgacaaactctatttcggc 1560
Db 95566 AGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGC 95625
Qy 1561 tttcgcggcggacgtttggatttaaacgggcattcgctttcgttccaccgtattcaaaat 1620
Db 95626 TTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAAT 95685
Qy 1621 accgatgaaggggcgatgattgncnatcataatgccacaacaacatccaccgttaccatt 1680
Db 95686 ACCGATGAAGGGGCGATGATTGTCAATCATAATGCCACAACAACATCCACCGTTACCATT 95745
Qy 1681 acagggaatgaaagtattacacaaccgagtggtaagaatatcaatagacttaattacagc 1740
Db 95746 ACAGGGAATGAAAGTATTACACAACCGAGTGGTAAGAATATCAATAGACTTAATTACAGC 95805
Qy 1741 aaagaaattgcctacaacggttggtttggcgagaaagatacgaccaaaacgaacgggcgg 1800
Db 95806 AAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGG 95865
Qy 1801 ctcaaccttgtttaccagcccgccgcagaagaccgcacccngctgctttccggcggaaca 1860
Db 95866 CTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACA 95925
Qy 1861 aatttaaacggcaacatcacgcaaacaaacggcaaactgtttttcagcggcagaccgaca 1920
Db 95926 AATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACA 95985
Qy 1921 ccgcacgcctacaatcatttaggaagcgggtggtcaaaaatggaaggtatcccacaagga 1980
Db 95986 CCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 96045
Qy 1981 gaaatcgtgtgggacaacgactggatcnaccgcacgtttaaagcggaaaatttccatatt 2040
Db 96046 GAAATCGTGTGGGACAACGACTGGATCAACCGCACGTTTAAAGCGGAAAATTTCCATATT 96105
Qy 2041 cagggcgggcaggcggtgatttcccgcaatgttgccaaagtggaaggcgattgncatttg 2100
Db 96106 CAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTG 96165
Qy 2101 agcaatcacgcccaagcagtttttggtgtcgcaccgcatcaaagccatacaatctgtaca 2160
Db 96166 AGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATCTGTACA 96225
Qy 2161 cgttcggactggacnggtctgacaaattgtgtcgaanaaancattaccgacgataaagtg 2220
Db 96226 CGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTG 96285
Qy 2221 attgcttcattgactaagacngacntnagcggcantgtnagnctnnccnatnacgntnnt 2280
Db 96286 ATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCAGCCTTGCCGATCACGCTCAT 96345
Qy 2281 tnaaanctcncngggcntgcnncactnaanggcaatcttagtgcaaatggcgatacacgt 2340
Db 96346 TTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGT 96405
Qy 2341 tatacagtcagccacaacgccacccaaaacggcaaccttagcctcgtgggcaatgcccaa 2400
Db 96406 TATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAA 96465
Qy 2401 gcaacatttaatcaagccacattaaacggcaacncatcggnttcgggcaatgcttcattt 2460
Db 96466 GCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTT 96525
Qy 2461 aatctaagcaacaacgccgcacaaaacggcagtctgacgctttccgacaacgctaaggca 2520
Db 96526 AATCTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCA 96585
Qy 2521 aacgtaagccattccgcactcaacggcaatgtctccctagccgataaggcagtattccat 2580

```
Db  96586 AACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT 96645

Qy   2581 tttgaaaacagccgctttaccggacaactcagcggcagcaaggana cagcattacactta 2640
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  96646 TTTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGATACAGCATTACACTTA 96705

Qy   2641 aaagacagcgaatggacgctgccgtcaggcacggaattaggcaatttaaaccttgacaac 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  96706 AAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAAC 96765

Qy   2701 gccaccattacactcaattccgcctatcgccacgatgctgcaggcgcgcaaaccggcagn 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  96766 GCCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGT 96825

Qy   2761 gtgtcagacacgccgcgccgccgttcgcgccgttccctattatccgttacaccgccaact 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  96826 GTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACT 96885

Qy   2821 tcggtagaatcccgtttcaacacgctgacggtaaacggcaaattgaacngtcaaggaaca 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  96886 TCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAAGGAACA 96945

Qy   2881 ttccgctttatgtcggaactcttcggctaccgaagcgacaaattgaagctggcggaaagt 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  96946 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGT 97005

Qy   2941 tccgaaggnacttacaccttggcggtcaacaataccggcaacgaacccgtaagcctcgat 3000
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97006 TCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGAT 97065

Qy   3001 caattgacggtagtggaagggaaagacaacaaaccgctgtccgaaaaccttaatttcacc 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97066 CAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACC 97125

Qy   3061 ctgcaaaacgaacacgtcgatgccggcgcgtggcgttaccaactcatccgcaaagacggc 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97126 CTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 97185

Qy   3121 gagttccgcctgcataatccggtcaaagaacaagagctttccgacaaactcggcaaggca 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97186 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCA 97245

Qy   3181 gaagccaaaaaacaggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcggcc 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97246 GAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCC 97305

Qy   3241 gggcgcgatgccgccgaaaagacagaaagcgttgccgaaccggcccggcnggcaggcggg 3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97306 GGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGG 97365

Qy   3301 gaaaatgtcggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaagac 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97366 GAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC 97425

Qy   3361 agcgcnttggcgaaacagcgcgaagcggaaacccggccggntaccaccgccttcccccgc 3420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97426 AGCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGC 97485

Qy   3421 gcccgcngcgcccgccgggatttgccgcaaccgcagccccaaccgcaacctcaaccccaa 3480
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97486 GCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAA 97545

Qy   3481 ccgcagcgcgacctgatnagccgttatgccaatagcggtttgagtgaattttccgccacg 3540
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  97546 CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACG 97605

Qy   3541 ctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagaccgccgc 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97606 CTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGC 97665

Qy   3601 aacgcngtttggacaagcngcatccggnacaccaaacactaccgttcgcaagatttccgc 3660
          ||||| |||||||||||| |||||||| ||||||||||||||||||||||||||||||||
Db  97666 AACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGC 97725

Qy   3661 gcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggcagcggg 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97726 GCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG 97785

Qy   3721 cgcgtcggcatcctgttttcgcacaaccggaccgaaaacancttcgacgacggcatcggc 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97786 CGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGC 97845

Qy   3781 aactcggcacggcttgcccacggcgccgttttcgggcaatacggcatcggcaggttcgac 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97846 AACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGAC 97905

Qy   3841 atcggcatcagcacgggcgcgggttttagcagcggcantctntcagacggcatcggaggc 3900
          |||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db  97906 ATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCAGTCTTTCAGACGGCATCGGAGGC 97965

Qy   3901 aaaatccgccgccgcgtgctgcattacggcattcaggcacgataccgcgccggtttcggc 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97966 AAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGC 98025

Qy   3961 ggattcggcatcgaaccgtacatcggcgcaacgcgctatttcgtccaaaaagcggattac 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98026 GGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTAC 98085

Qy   4021 cgctacgaaaacgtcaatatcgccacccccggtcttgcgttcaaccgntaccgngcgggc 4080
          |||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||
Db  98086 CGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGCTACCGCGCGGGC 98145

Qy   4081 attaaggcagattattcattcaaaccggcgcaacacatntccatcacnccttatttnagc 4140
          |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||| |||
Db  98146 ATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGC 98205

Qy   4141 ctgtcctataccgatgccgcttcgggcaaagtccgaacacgcgtcaataccgcngtattg 4200
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  98206 CTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTG 98265

Qy   4201 gctcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatcaaaggt 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98266 GCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGT 98325

Qy   4261 ttcacgctgtccntccacgctgccgccgccaaaggnccgcaactggaagcgcaacacagc 4320
          |||||||||||| |||||||||||||||||||||| ||||||||||||||||||||||||
Db  98326 TTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGC 98385

Qy   4321 gcgggcatcaaattaggctaccgctggtaa 4350
          |||||||||||||||||||||||||||||
Db  98386 GCGGGCATCAAATTAGGCTACCGCTGGTAA 98415
```

```
RESULT   3
NME296276
LOCUS       NME296276               4350 bp    DNA     linear   BCT 14-JAN-2002
DEFINITION  Neisseria meningitidis app gene for adhesion penetration protein,
            strain Z3515.
ACCESSION   AJ296276
VERSION     AJ296276.1  GI:11071862
KEYWORDS    adhesion penetration protein; APP gene.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (bases 1 to 4350)
  AUTHORS   Ala-Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2000) Ala-Aldeen D.A., Microbiology, University
            of Nottingham, Molecular Bacteriology and Immunology Group,
```

```
          University Hospital, Nottingham NG7 2UH, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .4350
                     /organism="Neisseria meningitidis"
                     /strain="Z3515"
                     /db_xref="taxon:487"
                     /country="Saudi Arabia"
                     /note="serogroup A"
     gene            1. .4350
                     /gene="app"
     CDS             1. .4350
                     /gene="app"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion penetration protein"
                     /protein_id="CAC14670.1"
                     /db_xref="GI:11071863"
                     /db_xref="SPTREMBL:Q9F3Z5"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHM
                     PRLHKFVTDAEPVEMTSDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGA
                     WLIGGNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLL
                     NGVLQTGYPYSGRENGFQLIRKDWFYDDIYRGDTHTVFFEPRSNGHFSFTSNNNGTGT
                     VTETNEKVSNPKLEVQTVRLFDESLNETDKEPVYAAGGVNQYRPRLNNGENLSFIDYG
                     NGKLILSNNINQGAGGLYFEGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRL
                     SKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVSGRGTVQLNA
                     DNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNATTTSTVTITGNESI
                     TQPSGKNINRLNYSKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLSGGTNLNG
                     NITQTNGKLFFSGRPTPHAYNHLGSGWSKMEGIPQGEIVWDNDWINRTFKAENFHIQG
                     GQAVISRNVAKVEGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEKTITDDKV
                     IASLTKTDISGNVSLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNGNLSLVGN
                     AQATFNQATLNGNTSASGNASFNLSNNAAQNGSLTLSDNAKANVSHSALNGNVSLADK
                     AVFHFENSRFTGQLSGSKDTALHLKDSEWTLPSGTELGNLNLDNATITLNSAYRHDAA
                     GAQTGSVSDTPRRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFRFMSELFGYRS
                     DKLKLAESSEGTYTLAVNNTGNEPVSLDQLTVVEGKDNKPLSENLNFTLQNEHVDAGA
                     WRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALIAAGRDAAEKT
                     ESVAEPARQAGGENVGIMQAEEEKKRVQADKDSALAKQREAETRPATTAFPRARRARR
                     DLPQPQPQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVW
                     TSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNS
                     ARLAHGAVFGQYGIGRFDIGISTGAGFSSGSLSDGIGGKIRRRVLHYGIQARYRAGFG
                     GFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPY
                     LSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGPQLE
                     AQHSAGIKLGYRW"
BASE COUNT      1236 a    1185 c    1073 g     856 t
ORIGIN
```

```
  Query Match             98.5%;  Score 4284.8;  DB 1;  Length 4350;
  Best Local Similarity   98.5%;  Pred. No. 0;
  Matches 4286;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAGCACCGTTTTTCTTACCAAATT 420

Qy     421 gtgaaaagaaataattataagcctgacaattcacacccttacaacggcgattancatatg 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     421 GTGAAAAGAAATAATTATAAGCCTGACAATTCACACCCTTACAACGGCGATTACCATATG 480

Qy     481 ccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgagg 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTCGAAATGACGAGTGACATGAGG 540

Qy     541 gggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacac 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGGAATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACAC 600

Qy     601 cactattggcgttatgatgatgacaaacacggcgatttatcctactccggcgcatggtta 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CACTATTGGCGTTATGATGATGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTA 660

Qy     661 attggcggcaatacacatatgcagggttggggaaataatggcgtanttagtttgagcggc 720
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     661 ATTGGCGGCAATACACATATGCAGGGTTGGGGAAATAATGGCGTAGTTAGTTTGAGCGGC 720

Qy     721 gatgtgcgccatgccaacgactatggccctatgccgattgcaggtgcggcaggcgacagc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGCAGGCGACAGC 780

Qy     781 ggttcgccaatgtttatttatgacaaaacaaacaataaatggctgctcaacggagtttta 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGTTCGCCAATGTTTATTTATGACAAAACAAACAATAAATGGCTGCTCAACGGAGTTTTA 840

Qy     841 caaaccggctacccttattccggcagggaaaacggtttccagctgatacgcaaagattgg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CAAACCGGCTACCCTTATTCCGGCAGGGAAAACGGTTTCCAGCTGATACGCAAAGATTGG 900

Qy     901 ttctacgatgacatttacagaggcgatacacataccgtctntttttgaaccgcgcagtaac 960
           ||||| ||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     901 TTCTATGATGACATTTACAGAGGCGATACACATACCGTCTTTTTTGAACCGCGCAGTAAC 960

Qy     961 ggacatttttccttttacatccaacaacaacggtacgggtacggtaacagaaaccaacgaa 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGACATTTTTCCTTTACATCCAACAACAACGGTACGGGTACGGTAACAGAAACCAACGAA 1020

Qy    1021 aaggtntccaatccaaagcttaaagtacagacagtccgactgtttgacgaatctttgaat 1080
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AAGGTTTCCAATCCAAAGCTTGAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAAT 1080

Qy    1081 gaaactgataaagaaccagtttacgcggcaggggtgttaatcagtaccgtccaaggtta 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GAAACTGATAAAGAACCAGTTTACGCGGCAGGGGGTGTTAATCAGTACCGTCCAAGGTTA 1140

Qy    1141 aacaacggtgaaaacctttctttttatcgattacggcaacggcaaactcatcttatcaaac 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AACAACGGTGAAAACCTTTCTTTTATCGATTACGGCAACGGCAAACTCATCTTATCAAAC 1200

Qy    1201 aacatcaaccaaggcgcgggcggtttgtattttgaaggtgattttacggtctcgcctgaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 AACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCCTGAA 1260

Qy    1261 aacaacgaaacgtggcaaggcgcgggcgttcatatcagtgaagacagtaccgttacttgg 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 AACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGG 1320

Qy    1321 aaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgtt 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 AAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTT 1380

Qy    1381 caagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtacagtcattttggat 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  1381 CAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGAT 1440
Qy  1441 cagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcggcttgntcagcggc 1500
Db  1441 CAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGC 1500
Qy  1501 aggggtacggtgcaactgaatgccgataatcagttcaaccccgacaaactctatttcggc 1560
Db  1501 AGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGC 1560
Qy  1561 tttcgcggcggacgtttggatttaaacgggcattcgctttcgttccaccgtattcaaaat 1620
Db  1561 TTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAAT 1620
Qy  1621 accgatgaaggggcgatgattgncnatcataatgccacaacaacatccaccgttaccatt 1680
Db  1621 ACCGATGAAGGGGCGATGATTGTCAATCATAATGCCACAACAACATCCACCGTTACCATT 1680
Qy  1681 acagggaatgaaagtattacacaaccgagtggtaagaatatcaatagacttaattacagc 1740
Db  1681 ACAGGGAATGAAAGTATTACACAACCGAGTGGTAAGAATATCAATAGACTTAATTACAGC 1740
Qy  1741 aaagaaattgcctacaacggttggtttggcgagaaagatacgaccaaaacgaacgggcgg 1800
Db  1741 AAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGG 1800
Qy  1801 ctcaaccttgtttaccagcccgccgcagaagaccgcacccngctgctttccggcggaaca 1860
Db  1801 CTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACA 1860
Qy  1861 aatttaaacggcaacatcacgcaaacaaacggcaaactgtttttcagcggcagaccgaca 1920
Db  1861 AATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACA 1920
Qy  1921 ccgcacgcctacaatcatttaggaagcgggtggtcaaaaatggaaggtatcccacaagga 1980
Db  1921 CCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 1980
Qy  1981 gaaatcgtgtgggacaacgactggatcnaccgcacgtttaaagcggaaaatttccatatt 2040
Db  1981 GAAATCGTGTGGGACAACGACTGGATCAACCGCACGTTTAAAGCGGAAAATTTCCATATT 2040
Qy  2041 cagggcgggcaggcggtgatttcccgcaatgttgccaaagtggaaggcgattgncatttg 2100
Db  2041 CAGGGCGGGCAGGCGGTGATTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTG 2100
Qy  2101 agcaatcacgcccaagcagtttttggtgtcgcaccgcatcaaagccatacaatctgtaca 2160
Db  2101 AGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCATACAATCTGTACA 2160
Qy  2161 cgttcggactggacnggtctgacaaattgtgtcgaanaaancattaccgacgataaagtg 2220
Db  2161 CGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTG 2220
Qy  2221 attgcttcattgactaagacngacntnagcggcantgtnagnctnnccnatnacgntnnt 2280
Db  2221 ATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCAGCCTTGCCGATCACGCTCAT 2280
Qy  2281 tnaaanctcncngggcntgcnncactnaanggcaatcttagtgcaaatggcgatacacgt 2340
Db  2281 TTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGT 2340
Qy  2341 tatacagtcagccacaacgccacccaaaacggcaaccttagcctcgtgggcaatgcccaa 2400
Db  2341 TATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAA 2400
Qy  2401 gcaacatttaatcaagccacattaaacggcaacncatcggnttcgggcaatgcttcattt 2460
Db  2401 GCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTT 2460
Qy  2461 aatctaagcaacaacgccgcacaaaacggcagtctgacgctttccgacaacgctaaggca 2520
Db  2461 AATCTAAGCAACAACGCCGCACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCA 2520
Qy  2521 aacgtaagccattccgcactcaacggcaatgtctccctagccgataaggcagtattccat 2580
Db  2521 AACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT 2580
Qy  2581 tttgaaaacagccgctttaccggacaactcagcggcagcaaggganacagcattacactta 2640
Db  2581 TTTGAAAACAGCCGCTTTACCGGACAACTCAGCGGCAGCAAGGATACAGCATTACACTTA 2640
Qy  2641 aaagacagcgaatggacgctgccgtcaggcacggaattaggcaatttaaaccttgacaac 2700
Db  2641 AAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAAC 2700
Qy  2701 gccaccattacactcaattccgcctatcgccacgatgctgcaggcgcgcaaaccggcagn 2760
Db  2701 GCCACCATTACACTCAATTCCGCCTATCGCCACGATGCTGCAGGCGCGCAAACCGGCAGT 2760
Qy  2761 gtgtcagacacgccgcgccgccgttcgcgccgttccctattatccgttacaccgccaact 2820
Db  2761 GTGTCAGACACGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACT 2820
Qy  2821 tcggtagaatcccgtttcaacacgctgacggtaaacggcaaattgaacngtcaaggaaca 2880
Db  2821 TCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAAGGAACA 2880
Qy  2881 ttccgctttatgtcggaactcttcggctaccgaagcgacaaattgaagctggcggaaagt 2940
Db  2881 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTGAAGCTGGCGGAAAGT 2940
Qy  2941 tccgaaggnacttacaccttggcggtcaacaataccggcaacgaacccgtaagcctcgat 3000
Db  2941 TCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCCGTAAGCCTCGAT 3000
Qy  3001 caattgacggtagtggaagggaaagacaacaaaccgctgtccgaaaaccttaatttcacc 3060
Db  3001 CAATTGACGGTAGTGGAAGGGAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACC 3060
Qy  3061 ctgcaaaacgaacacgtcgatgccggcgcgtggcgttaccaactcatccgcaaagacggc 3120
Db  3061 CTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3120
Qy  3121 gagttccgcctgcataatccggtcaaagaacaagagctttccgacaaactcggcaaggca 3180
Db  3121 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCA 3180
Qy  3181 gaagccaaaaaacaggcggaaaaagacaacgcgcaaagccttgacgcgctgattgcggcc 3240
Db  3181 GAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCC 3240
Qy  3241 gggcgcgatgccgccgaaaagacagaaagcgttgccgaaccggcccggcnggcaggcggg 3300
Db  3241 GGGCGCGATGCCGCCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGG 3300
Qy  3301 gaaaatgtcggcattatgcaggcggaggaagagaaaaaacgggtgcaggcggataaagac 3360
Db  3301 GAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC 3360
Qy  3361 agcgcnttggcgaaacagcgcgaagcggaaacccggccggntaccaccgccttcccccgc 3420
Db  3361 AGCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGC 3420
Qy  3421 gcccgcngcgcccgccgggatttgccgcaaccgcagccccaaccgcaacctcaaccccaa 3480
Db  3421 GCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCTCAACCCCAA 3480
Qy  3481 ccgcagcgcgacctgatnagccgttatgccaatagcggtttgagtgaattttccgccacg 3540
Db  3481 CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACG 3540
Qy  3541 ctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaagaccgccgc 3600
Db  3541 CTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGC 3600
```

```
Qy   3601 aacgcngtttggacaagcngcatccggnacaccaaacactaccgttcgcaagatttccgc 3660
          ||||| ||||||||||| |||||||| ||||||||||||||||||||||||||||||||||
Db   3601 AACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGC 3660

Qy   3661 gcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctcggcagcggg 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG 3720

Qy   3721 cgcgtcggcatcctgttttcgcacaaccggaccgaaaacancttcgacgacggcatcggc 3780
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   3721 CGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGC 3780

Qy   3781 aactcggcacggcttgcccacggcgccgttttcgggcaatacggcatcggcaggttcgac 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGAC 3840

Qy   3841 atcggcatcagcacgggcgcgggttttagcagcggcantctntcagacggcatcggaggc 3900
          |||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db   3841 ATCGGCATCAGCACGGGCGCGGGTTTTAGCAGCGGCAGTCTTTCAGACGGCATCGGAGGC 3900

Qy   3901 aaaatccgccgccgcgtgctgcattacggcattcaggcacgataccgcgccggtttcggc 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 AAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGC 3960

Qy   3961 ggattcggcatcgaaccgtacatcggcgcaacgcgctatttcgtccaaaaagcggattac 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GGATTCGGCATCGAACCGTACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTAC 4020

Qy   4021 cgctacgaaaacgtcaatatcgccacccccggtcttgcgttcaaccgntaccgngcgggc 4080
          |||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||
Db   4021 CGCTACGAAAACGTCAATATCGCCACCCCCGGTCTTGCGTTCAACCGCTACCGCGCGGGC 4080

Qy   4081 attaaggcagattattcattcaaaccggcgcaacacatntccatcacnccttatttnagc 4140
          |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||| |||
Db   4081 ATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGC 4140

Qy   4141 ctgtcctataccgatgccgcttcgggcaaagtccgaacacgcgtcaataccgcngtattg 4200
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   4141 CTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTG 4200

Qy   4201 gctcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaaatcaaaggt 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 GCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGT 4260

Qy   4261 ttcacgctgtccntccacgctgccgccgccaaaggnccgcaactggaagcgcaacacagc 4320
          |||||||||||| |||||||||||||||||||||| ||||||||||||||||||||||||
Db   4261 TTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGC 4320

Qy   4321 gcgggcatcaaattaggctaccgctggtaa 4350
          |||||||||||||||||||||||||||||
Db   4321 GCGGGCATCAAATTAGGCTACCGCTGGTAA 4350


RESULT   4
A96616
LOCUS       A96616                 4374 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 649 from Patent WO9924578.
ACCESSION   A96616
VERSION     A96616.1  GI:6780200
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4374)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 649 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="unidentified"
                     /db_xref="taxon:32644"

BASE COUNT     1266 a   1163 c   1066 g    879 t
ORIGIN


  Query Match             80.6%;  Score 3508;  DB 6;  Length 4374;
  Best Local Similarity   88.4%;  Pred. No. 0;
  Matches 3876;  Conservative    0;  Mismatches  462;  Indels   48;  Gaps    8;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy    361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
          |||||||||||||||||||||||| |||||||||| || |||||| |||| |||||
Db    361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy    421 gtgaaaagaaataattataagcctgacaattca---caccttacaacggcgattancat 477
          |||||| | |||||||||| | | | | |  || |||||  ||||||||| |||
Db    421 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 480

Qy    478 atgccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatg 537
          |||||||||||||||||||||||||||||||||||||||| |||||||| ||| | |||
Db    481 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 540

Qy    538 agggggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcagga 597
             |||   |  |||  |||| || | || || || || ||||| || || ||  ||||
Db    541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600

Qy    598 caccactattggcgttatgatgatga------caaacacggcgatttatcctactccggc 651
            || |||||||| | |||||| ||       |||  || ||||    || ||
Db    601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660

Qy    652 gca------------tggttaattggcggcaatacacatatgcagggttggggaaataat 699
          |||             ||| |  |||| ||||||||    |   ||   | |   |  | |
Db    661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

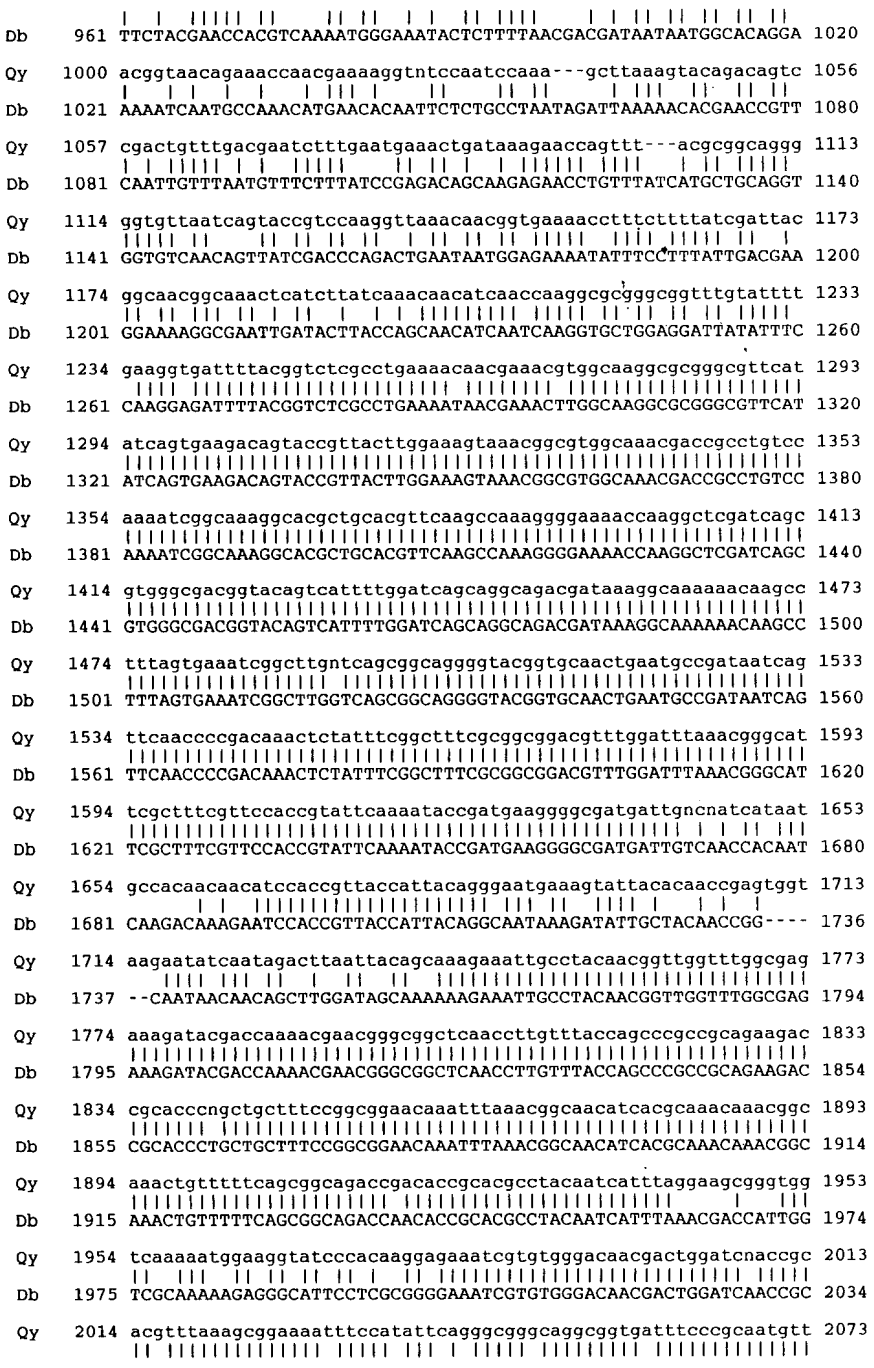
Qy    700 ggcgtanttagtttgagcggcgatgtgcgccatgccaacgactatggccctatgccgatt 759
          |||  | | |  ||  | | ||         |   | |  |||||   | | || |
Db    721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780

Qy    760 gcaggtgcggcaggcgacagcggttcgccaatgtttatttatgacaaaacaaacaataaa 819
          | |||  |    |||||||| || || ||||||||||| |||||     |||  | ||
Db    781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840

Qy    820 tggctgctcaacggagttttacaaaccggctacccttattccggcagggaaaacggtttc 879
          ||| |  | || || || || ||||| ||| |||| |||   || |     || || |||
Db    841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

Qy    880 cagctgatacgcaaagattggttctacgatgacatttacagaggcgatacacataccgtc 939
          |||||| | || |||||||||||||| ||||| || |     || ||||| ||| | ||
Db    901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960

Qy    940 tnttttgaaccgcgcagtaacggacatttttcctttacatccaacaacaacggtacgggt 999
```


Db 961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 1020
Qy 1000 acggtaacagaaaccaacgaaaaggtntccaatccaaa---gcttaaagtacagacagtc 1056
Db 1021 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080
Qy 1057 cgactgtttgacgaatctttgaatgaaactgataaagaaccagttt---acgcggcaggg 1113
Db 1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140
Qy 1114 ggtgttaatcagtaccgtccaaggttaaacaacggtgaaaaccttttcttttatcgattac 1173
Db 1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200
Qy 1174 ggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcgggcggtttgtatttt 1233
Db 1201 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 1260
Qy 1234 gaaggtgattttacggtctcgcctgaaaacaacgaaacgtggcaaggcgcgggcgttcat 1293
Db 1261 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320
Qy 1294 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1353
Db 1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380
Qy 1354 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1413
Db 1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440
Qy 1414 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1473
Db 1441 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500
Qy 1474 tttagtgaaatcggcttgntcagcggcaggggtacggtgcaactgaatgccgataatcag 1533
Db 1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560
Qy 1534 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1593
Db 1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1620
Qy 1594 tcgctttcgttccaccgtattcaaaatacCgatgaaggggcgatgattgncnatcataat 1653
Db 1621 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
Qy 1654 gccacaacaacatccaccgttaccattacagggaatgaaagtattacacaaccgagtggt 1713
Db 1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGG---- 1736
Qy 1714 aagaatatcaatagacttaattacagcaaagaaattgcctacaacggttggtttggcgag 1773
Db 1737 --CAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1794
Qy 1774 aaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagac 1833
Db 1795 AAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGAC 1854
Qy 1834 cgcacccngctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggc 1893
Db 1855 CGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGC 1914
Qy 1894 aaactgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtgg 1953
Db 1915 AAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGG 1974
Qy 1954 tcaaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgactggatcnaccgc 2013
Db 1975 TCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGC 2034
Qy 2014 acgtttaaagcggaaaatttccatattcagggcgggcaggcggtgatttcccgcaatgtt 2073
Db 2035 ACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTT 2094
Qy 2074 gccaaagtggaaggcgattgncatttgagcaatcacgcccaagcagtttttggtgtcgca 2133
Db 2095 GCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCA 2154
Qy 2134 ccgcatcaaagccatacaatctgtacacgttcggactggacnggtctgacaaattgtgtc 2193
Db 2155 CCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 2214
Qy 2194 gaanaaancattaccgacgataaagtgattgcttcattgactaagacngacntnagcggc 2253
Db 2215 GAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGC 2274
Qy 2254 antgtnagnctnnccnatnacgntnnttnaaanctcncngggcntgcnncactnaanggc 2313
Db 2275 AATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 2334
Qy 2314 aatcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggc 2373
Db 2335 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGC 2394
Qy 2374 aaccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2433
Db 2395 AACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 2454
Qy 2434 ncatcggnttcgggcaatgcttcatttaatctaagcaacaacgccgcacaaaacggcagt 2493
Db 2455 ACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGT 2514
Qy 2494 ctgacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtc 2553
Db 2515 CTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTC 2574
Qy 2554 tccctagccgataaggcagtattccattttgaaaacagccgctttaccggacaactcagc 2613
Db 2575 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGC 2634
Qy 2614 ggcagcaaggancacgcattacacttaaaagacagcgaatggacgctgccgtcaggcacg 2673
Db 2635 GGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACG 2694
Qy 2674 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccac 2733
Db 2695 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2754
Qy 2734 gatgctgcaggcgcgcaaaccggcagngtgtcagacacgccgcgccgccgttcg------ 2787
Db 2755 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814
Qy 2788 ---cgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacg 2844
Db 2815 TCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 2874
Qy 2845 ctgacggtaaacggcaaattgaacngtcaaggaacattccgctttatgtcggaactcttc 2904
Db 2875 CTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934
Qy 2905 ggctaccgaagcgacaaattgaagctggcggaaagttccgaaggnacttacaccttggcg 2964
Db 2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCG 2994
Qy 2965 gtcaacaataccggcaacgaacccgtaagcctcgatcaattgacggtagtggaagggaaa 3024
Db 2995 GTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
Qy 3025 gacaacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgcc 3084
Db 3055 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCC 3114
Qy 3085 ggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtc 3144
Db 3115 GGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC 3174

```
Qy   3145 aaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaa 3204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3175 AAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 3234

Qy   3205 gacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgccgaaaagaca 3264
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   3235 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACA 3294

Qy   3265 gaaagcgttgccgaaccggcccggcnggcaggcggggaaaatgtcggcattatgcaggcg 3324
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   3295 GAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3354

Qy   3325 gaggaagagaaaaaacgggtgcaggcggataaagacagcgcnttggcgaaacagcgcgaa 3384
          |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db   3355 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAA 3414

Qy   3385 gcggaaacccggccggntaccaccgccttccccgcgcccgcngcgcccgccgggatttg 3444
          |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db   3415 GCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTG 3474

Qy   3445 ccgcaaccgcagccccaaccgcaacctcaaccccaaccgcagcgcgacctgatnagccgt 3504
          ||||||| ||| |||||||||      || ||||||||||||||||||||||| ||||||
Db   3475 CCGCAACTGCAACCCCAACCG------CAGCCCCAACCGCAGCGCGACCTGATCAGCCGT 3528

Qy   3505 tatgccaatagcggtttgagtgaatttttccgccacgctcaacagcgttttcgccgtacag 3564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3529 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAG 3588

Qy   3565 gacgaattggaccgcgtgtttgccgaagaccgccgcaacgcngtttggacaagcngcatc 3624
          |||||||| |||||||| |||||||||||||||||||||||| |||||||||||| |||||
Db   3589 GACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATC 3648

Qy   3625 cggnacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctg 3684
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3649 CGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 3708

Qy   3685 cgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcac 3744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3709 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCAC 3768

Qy   3745 aaccggaccgaaaacancttcgacgacggcatcggcaactcggcacggcttgcccacggc 3804
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   3769 AACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3828

Qy   3805 gccgttttcgggcaatacggcatcggcaggttcgacatcggcatcagcacgggcgcgggt 3864
          |||||||||||||||||||||||| ||||||| |||||||||||| ||||||||||||||
Db   3829 GCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGT 3888

Qy   3865 tttagcagcggcantctntcagacggcatcggaggcaaaatccgccgccgcgtgctgcat 3924
          |||||||||||||  || ||||||||||||||||||||||||||||||||||||||||||||
Db   3889 TTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCAT 3948

Qy   3925 tacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgtacatc 3984
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   3949 TACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATC 4008

Qy   3985 ggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgcc 4044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4009 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCC 4068

Qy   4045 acccccggtcttgcgttcaaccgntaccgngcgggcattaaggcagattattcattcaaa 4104
          |||||||| |||| |||||||| |||| ||||||||||||||||||||||||||||||||
Db   4069 ACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAA 4128

Qy   4105 ccggcgcaacacatntccatcacnccttatttnagcctgtcctataccgatgccgcttcg 4164
          ||||||||||||||| |||||||| |||||||| ||||||||||||||||||||||||||||
Db   4129 CCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCG 4188

Qy   4165 ggcaaagtccgaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgc 4224
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   4189 GGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGC 4248

Qy   4225 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgcc 4284
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   4249 AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCC 4308

Qy   4285 gccgccaaaggnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4344
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4309 GCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGC 4368

Qy   4345 tggtaa 4350
          ||||||
Db   4369 TGGTAA 4374


RESULT   5
AX044007
LOCUS       AX044007               4374 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 86 from Patent WO0066791.
ACCESSION   AX044007
VERSION     AX044007.1  GI:11342893
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4374)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 86 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
BASE COUNT     1266 a   1163 c   1066 g    879 t
ORIGIN


  Query Match             80.6%;  Score 3508;  DB 6;  Length 4374;
  Best Local Similarity   88.4%;  Pred. No. 0;
  Matches 3876;  Conservative    0;  Mismatches  462;  Indels   48;  Gaps    8;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

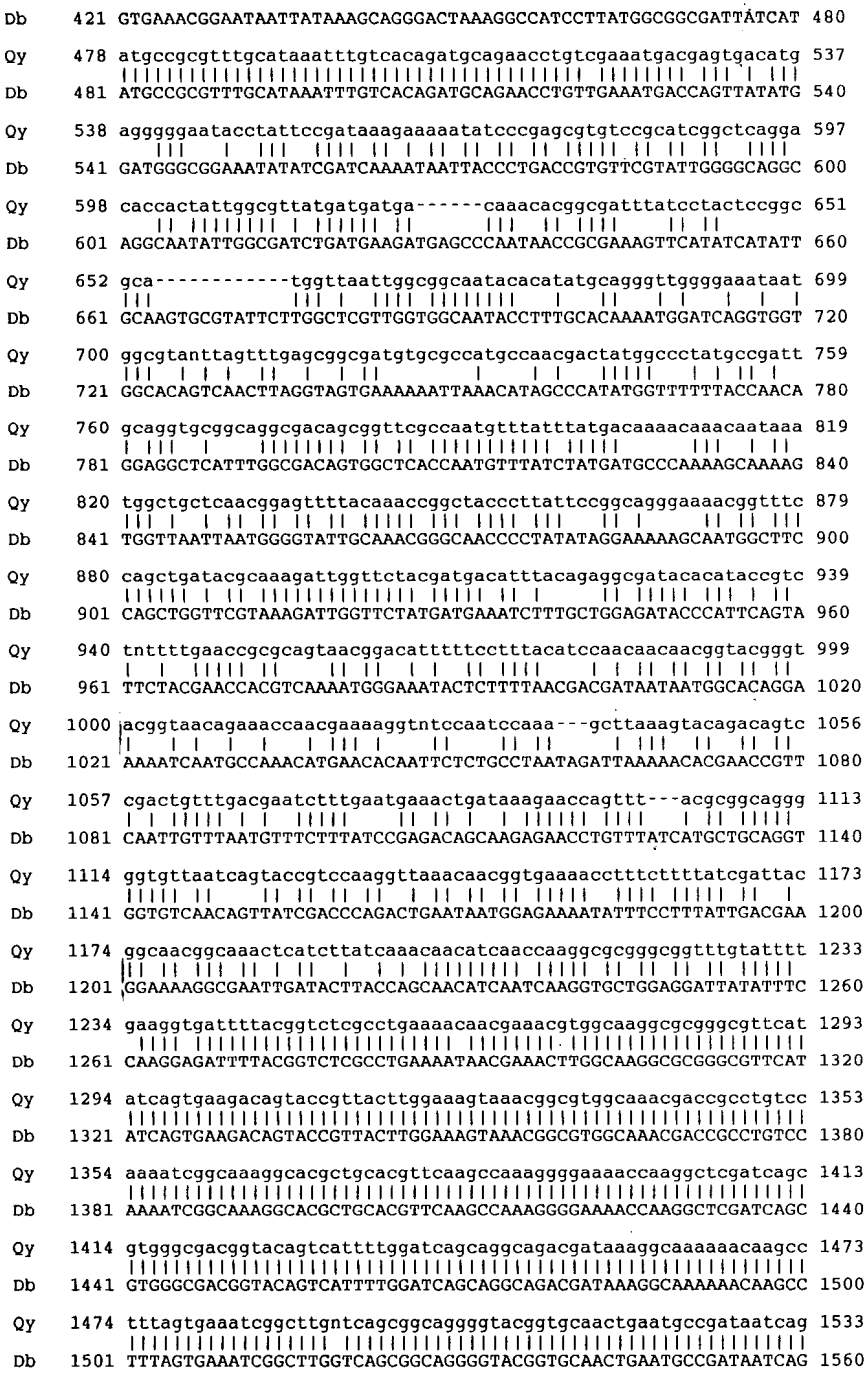
Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

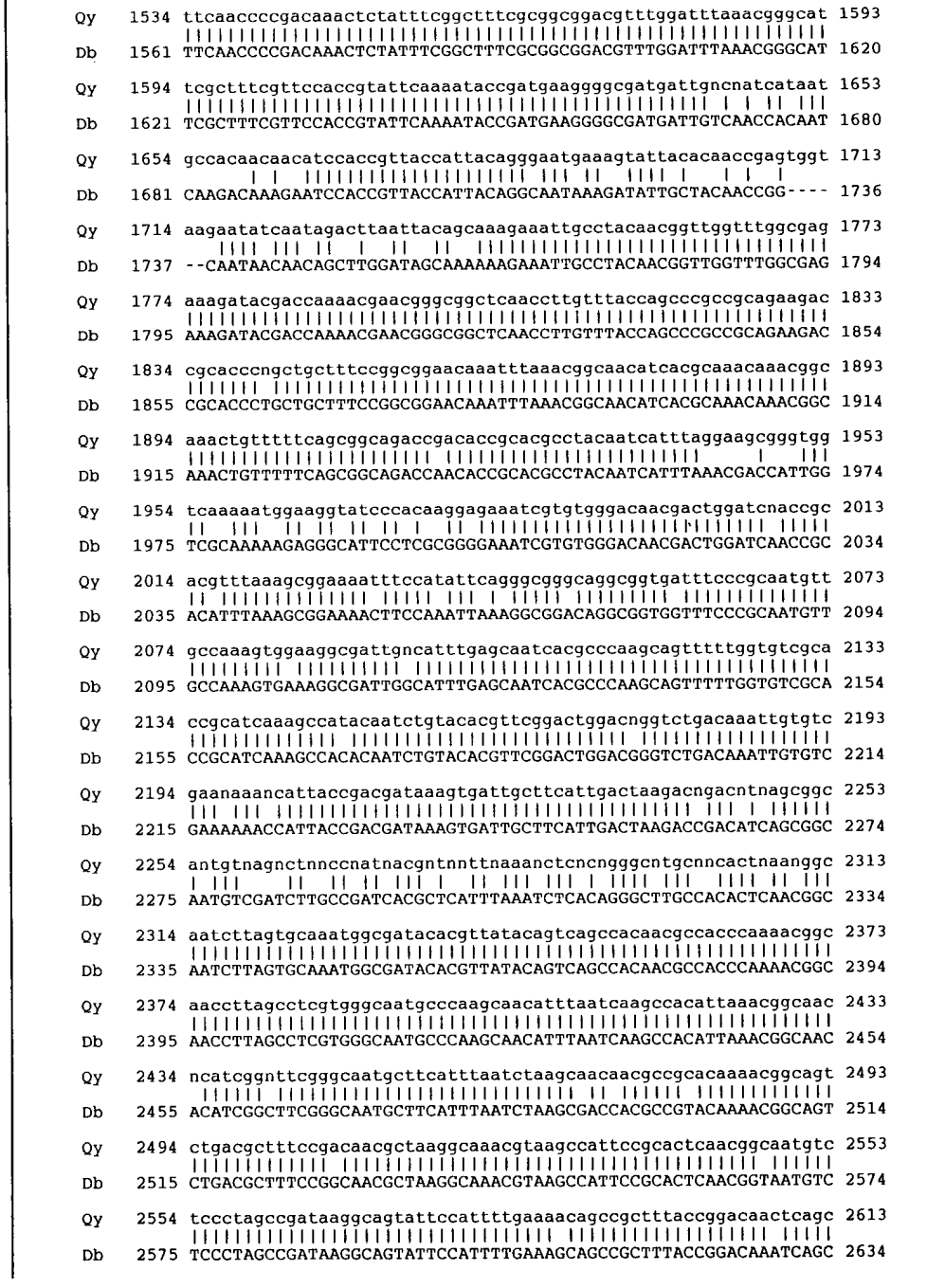
Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy    361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
          |||||||||||||||||||||||||| |||||||||| || |||||| |||| ||||||
Db    361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy    421 gtgaaaagaaataattataagcctgacaattca---caccttacaacggcgattancat 477
          |||||| | |||||||||||  | |  | |  |   || ||||||   |||||||||| |||
```





```
Qy   2614 ggcagcaaggana cagcattacacttaaaagacagcgaatggacgctgccgtcaggcacg 2673
          ||| ||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db   2635 GGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACG 2694

Qy   2674 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccac 2733
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2695 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2754

Qy   2734 gatgctgcaggcgcgcaaaccggcagngtgtcagacacgccgcgccgccgttcg------ 2787
          ||||| ||||| |||||||||||||| | | |||| |||||||||||||||||
Db   2755 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814

Qy   2788 ---cgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacg 2844
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2815 TCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 2874

Qy   2845 ctgacggtaaacggcaaattgaacngtcaaggaacattccgctttatgtcggaactcttc 2904
          |||||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||
Db   2875 CTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934

Qy   2905 ggctaccgaagcgacaaattgaagctggcggaaagttccgaaggnacttacaccttggcg 2964
          ||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCG 2994

Qy   2965 gtcaacaataccggcaacgaacccgtaagcctcgatcaattgacggtagtggaagggaaa 3024
          |||||||||||||||||||||||| | |||||||||.|||||||||||||||||||||| |||
Db   2995 GTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054

Qy   3025 gacaacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgcc 3084
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3055 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCC 3114

Qy   3085 ggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtc 3144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3115 GGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC 3174

Qy   3145 aaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaa 3204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3175 AAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 3234

Qy   3205 gacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgccgaaaagaca 3264
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   3235 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACA 3294

Qy   3265 gaaagcgttgccgaaccggcccggcnggcaggcggggaaaatgtcggcattatgcaggcg 3324
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   3295 GAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3354

Qy   3325 gaggaagagaaaaaacgggtgcaggcggataaagacagcgcnttggcgaaacagcgcgaa 3384
          |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||
Db   3355 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAA 3414

Qy   3385 gcggaaacccggccggntaccaccgccttcccccgcgcccgcngcgcccgccgggatttg 3444
          |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db   3415 GCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTG 3474

Qy   3445 ccgcaaccgcagccccaaccgcaacctcaaccccaaccgcagcgcgacctgatnagccgt 3504
          ||||||| ||| |||||||||      || ||||||||||||||||||||||| ||||||
Db   3475 CCGCAACTGCAACCCCAACCG------CAGCCCCAACCGCAGCGCGACCTGATCAGCCGT 3528

Qy   3505 tatgccaatagcggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacag 3564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3529 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAG 3588

Qy   3565 gacgaattggaccgcgtgtttgccgaagaccgccgcaacgcngtttggacaagcngcatc 3624
          |||||||| |||||||| |||||||||||||||||||||| |||||||||||| |||||
Db   3589 GACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATC 3648

Qy   3625 cggnacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctg 3684
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3649 CGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 3708

Qy   3685 cgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcac 3744
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3709 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCAC 3768

Qy   3745 aaccggaccgaaaacancttcgacgacggcatcggcaactcggcacggcttgcccacggc 3804
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3769 AACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3828

Qy   3805 gccgttttcgggcaatacggcatcggcaggttcgacatcggcatcagcacgggcgcgggt 3864
          |||||||||||||||||||||||||| ||||||| |||||||||||||| ||||||||||
Db   3829 GCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGT 3888

Qy   3865 tttagcagcggcantctntcagacggcatcggaggcaaaatccgccgccgcgtgctgcat 3924
          |||||||||||||  || ||||||||||||||||||||||||||||||||||||||||||||
Db   3889 TTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCAT 3948

Qy   3925 tacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgtacatc 3984
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   3949 TACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATC 4008

Qy   3985 ggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgcc 4044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4009 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCC 4068

Qy   4045 acccccggtcttgcgttcaaccgntaccgngcgggcattaaggcagattattcattcaaa 4104
          |||||||| |||||| |||||||| |||||| ||||||||||||||||||||||||||||
Db   4069 ACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAA 4128

Qy   4105 ccggcgcaacacatntccatcacnccttatttnagcctgtcctataccgatgccgcttcg 4164
          |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db   4129 CCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCG 4188

Qy   4165 ggcaaagtccgaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgc 4224
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   4189 GGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGC 4248

Qy   4225 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgcc 4284
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   4249 AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCC 4308

Qy   4285 gccgccaaaggnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4344
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4309 GCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGC 4368

Qy   4345 tggtaa 4350
          ||||||
Db   4369 TGGTAA 4374


RESULT   6
AE002547/c
LOCUS       AE002547               10074 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 189 of 206
            of the complete genome.
ACCESSION   AE002547 AE002098
VERSION     AE002547.1  GI:7227244
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10074)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
```

```
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10074)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10074
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            704. .1375
                     /gene="NMB1983"
     CDS             704. .1375
                     /gene="NMB1983"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF42311.1"
                     /db_xref="GI:7227245"
                     /translation="MFPDKYKLSLKENIFLAKKVLVAQIHNLSRFENCQTTLLQTEQI
                     IHGKNVASASLEDIQTILNLKRAYQYVISHISNGEPVDISLLKKINNIVAKDDSLAPG
                     DFRTGSVGVTLLDGSRHAPNPVKEIEVARVLQNIGLQSGSTTEAAVRFMLYCMRQQVF
                     WDGNKRTATLFANGLMMAGGCGILEISEMQMPQFNEKLSAFYRSGDDTDISKFVYQNC
                     ISGID"
     gene            1437. .2548
                     /gene="NMB1984"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     SP:Q00840 GB:Z11857 PID:45066 percent identity: 97.40;
                     identified by sequence similarity; putative;IS1106
                     transposase, authentic frameshift"
     gene            complement(2864. .7237)
                     /gene="NMB1985"
     CDS             complement(2864. .7237)
                     /gene="NMB1985"
                     /note="similar to GP:1408569 percent identity: 76.08;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion and penetration protein"
                     /protein_id="AAF42312.1"
                     /db_xref="GI:7227246"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVFYEPRQNGKYSFND
                     DNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINQGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNG
                     NLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHDAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFR
                     FMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFT
                     LQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALI
                     AAGRDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATT
                     AFPRARRARRDLPQLQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFA
                     EDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENT
                     FDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVLHYGIQ
                     ARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA
                     QHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAA
                     AAKGPQLEAQHSAGIKLGYRW"
     gene            complement(7371. .7802)
                     /gene="NMB1986"
     CDS             complement(7371. .7802)
                     /gene="NMB1986"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF42313.1"
                     /db_xref="GI:7227247"
                     /translation="MTENAQDKARQAVETVVKSPELVEQILSDEYVQIMIARCFHSGP
                     LPPPSDLAQYNDIISNGADRIMAMAEKEQAVRHETIRQDQTFNRRGQLYGFISAILIL
                     LFAVFLVWSGYPATAASLAGGTVFALAGAFVIGRSRDQGKN"
     gene            8087. .9433
                     /gene="NMB1987"
     CDS             8087. .9433
                     /gene="NMB1987"
                     /note="similar to GB:L10328 SP:P25522 PID:290554 GB:U00096
                     PID:2367268 percent identity: 69.20; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="thiophene and furan oxidation protein ThdF"
                     /protein_id="AAF42314.1"
                     /db_xref="GI:7227248"
                     /translation="MSDNVPTIAAVATAPGRGGVGVIRISGKNLLPMAQALCGKTPKP
                     RTATYADFTDTDGQAIDSGLLLFFAAPASFTGEDVIELQGHGGPVVMDMLLNRCLELG
                     ARLAEPGEFTKRAFLNDKLDLAQAEGVADLIDASSRSAARLALRSLKGDFSRRIHGLV
                     EDLITLRMLVEATLDFPEEDIDFLEAADARGKLDGLRRAVDDVLANAQQGAILREGLN
                     VVLVGAPNVGKSSLLNALAGDEVAIVTDIAGTTRDAVRERILIDGVPVHIVDTAGLRE
                     TDDVVERIGIERSRKAVSEADVALVLVDPREGLNEKTRAILDALPPELKRIEIHSKSD
                     LHAHAAGGFGTGAETVIALSAKTGDGLDALKRTLLREAGWQGESEGLFLARTRHVNAL
                     KAAQEELSLAALCGNHQIELFAEHLRLAQVACGEITGEFTADDLLGVIFSRFCIGK"
BASE COUNT     2307 a    2467 c    2721 g    2579 t
ORIGIN


  Query Match             80.6%;  Score 3508;  DB 1;  Length 10074;
  Best Local Similarity   88.4%;  Pred. No. 0;
  Matches 3876;  Conservative    0;  Mismatches  462;  Indels   48;  Gaps    8;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7237 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 7178

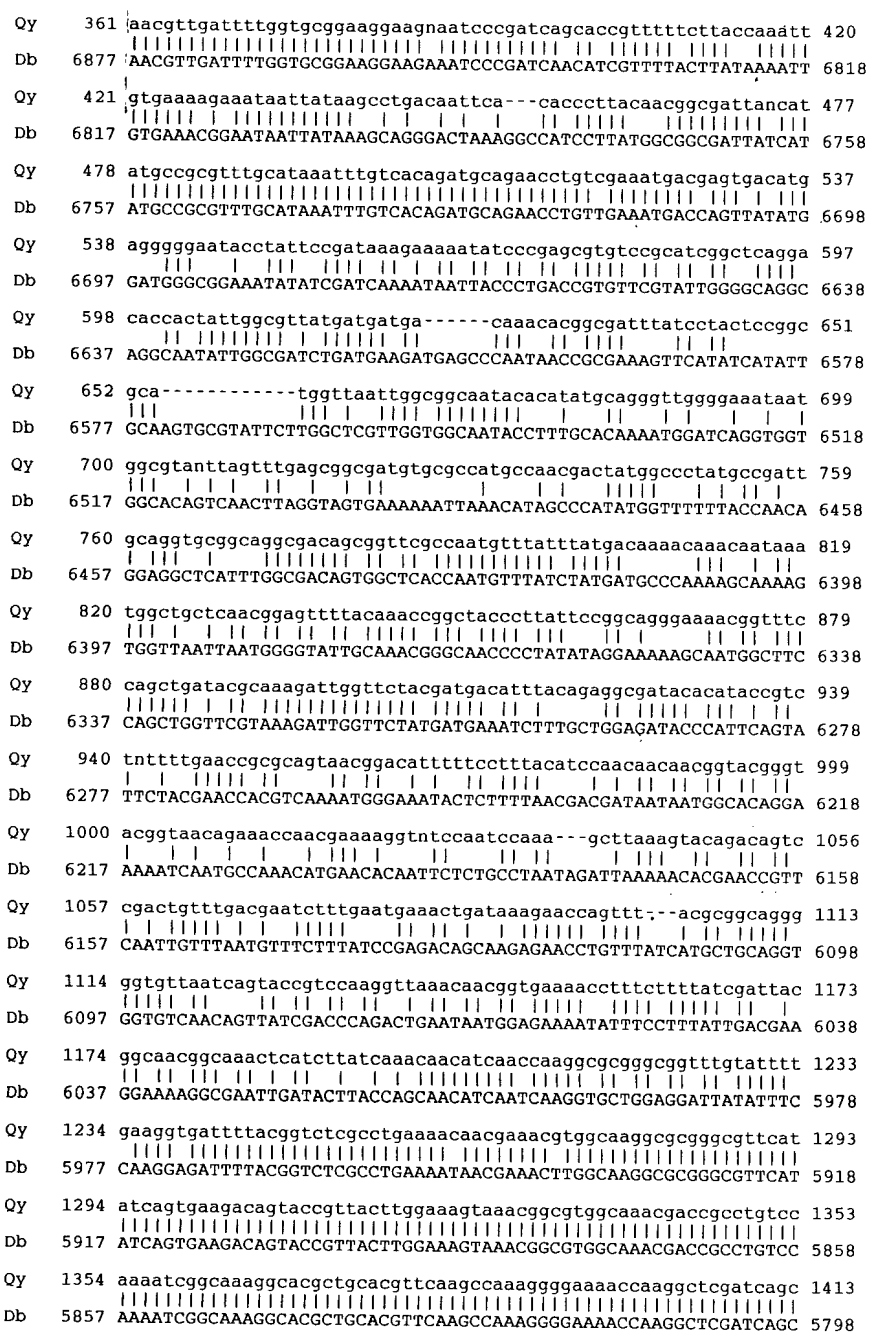
Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7177 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 7118

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7117 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 7058

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    7057 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 6998

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6997 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 6938

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6937 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 6878
```


Qy 361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
Db 6877 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 6818
Qy 421 gtgaaaagaaataattataagcctgacaattca---cacccttacaacggcgattancat 477
Db 6817 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 6758
Qy 478 atgccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatg 537
Db 6757 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 6698
Qy 538 aggggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcagga 597
Db 6697 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 6638
Qy 598 caccactattggcgttatgatgatga------caaacacggcgatttatcctactccggc 651
Db 6637 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 6578
Qy 652 gca----------tggttaattggcggcaatacacatatgcagggttggggaaataat 699
Db 6577 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 6518
Qy 700 ggcgtanttagtttgagcggcgatgtgcgccatgccaacgactatggccctatgccgatt 759
Db 6517 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 6458
Qy 760 gcaggtgcggcaggcgacagcggttcgccaatgtttatttatgacaaaacaaacaataaa 819
Db 6457 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 6398
Qy 820 tggctgctcaacggagttttacaaaccggctacccttattccggcagggaaaacggtttc 879
Db 6397 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 6338
Qy 880 cagctgatacgcaaagattggttctacgatgacatttacagaggcgatacacatacgtc 939
Db 6337 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 6278
Qy 940 tnttttgaaccgcgcagtaacggacatttttccttacatccaacaacaacggtacgggt 999
Db 6277 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 6218
Qy 1000 acggtaacagaaaccaacgaaaaggtntccaatccaaa---gcttaaagtacagacagtc 1056
Db 6217 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 6158
Qy 1057 cgactgtttgacgaatctttgaatgaaactgataaagaaccagttt---acgcggcaggg 1113
Db 6157 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 6098
Qy 1114 ggtgttaatcagtaccgtccaaggttaaacaacggtgaaaacctttctttttatcgattac 1173
Db 6097 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 6038
Qy 1174 ggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcgggcggtttgtatttt 1233
Db 6037 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 5978
Qy 1234 gaaggtgattttacggtctcgcctgaaaacaacgaaacgtggcaaggcgcgggcgttcat 1293
Db 5977 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 5918
Qy 1294 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1353
Db 5917 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 5858
Qy 1354 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1413
Db 5857 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 5798
Qy 1414 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1473
Db 5797 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 5738
Qy 1474 tttagtgaaatcggcttgntcagcggcaggggtacggtgcaactgaatgccgataatcag 1533
Db 5737 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 5678
Qy 1534 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1593
Db 5677 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 5618
Qy 1594 tcgctttcgttccaccgtattcaaaatacccgatgaaggggcgatgattgncnatcataat 1653
Db 5617 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 5558
Qy 1654 gccacaacaacatccaccgttaccattacagggaatgaaagtattacacaaccgagtggt 1713
Db 5557 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGG---- 5500
Qy 1714 aagaatatcaatagacttaattacagcaaagaaattgcctacaacggttggtttggcgag 1773
Db 5501 --CAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 5444
Qy 1774 aaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagac 1833
Db 5443 AAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGAC 5384
Qy 1834 cgcacccngctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggc 1893
Db 5383 CGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGC 5324
Qy 1894 aaactgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtgg 1953
Db 5323 AAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGG 5264
Qy 1954 tcaaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgactggatcnaccgc 2013
Db 5263 TCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGC 5204
Qy 2014 acgtttaaagcggaaaatttccatattcagggcgggcaggcggtgatttcccgcaatgtt 2073
Db 5203 ACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTT 5144
Qy 2074 gccaaagtggaaggcgattgncatttgagcaatcacgcccaagcagtttttggtgtcgca 2133
Db 5143 GCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCA 5084
Qy 2134 ccgcatcaaagccatacaatctgtacacgttcggactggacnggtctgacaaattgtgtc 2193
Db 5083 CCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 5024
Qy 2194 gaanaaancattaccgacgataaagtgattgcttcattgactaagacngacntnagcggc 2253
Db 5023 GAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGC 4964
Qy 2254 antgtnagnctnnccnatnacgntnnttnaaanctcncngggcntgcnncactnaanggc 2313
Db 4963 AATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 4904
Qy 2314 aatcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggc 2373
Db 4903 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGC 4844
Qy 2374 aaccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2433
Db 4843 AACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 4784
Qy 2434 ncatcggnttcgggcaatgcttcatttaatctaagcaacaacgccgcacaaaacggcagt 2493
Db 4783 ACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGT 4724
Qy 2494 ctgacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtc 2553

```
Db   4723 CTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTC 4664

Qy   2554 tccctagccgataaggcagtattccatttttgaaaacagccgctttaccggacaactcagc 2613
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4663 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGC 4604

Qy   2614 ggcagcaaggaaacagcattacacttaaaagacagcgaatggacgctgccgtcaggcacg 2673
          ||| ||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db   4603 GGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACG 4544

Qy   2674 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccac 2733
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4543 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 4484

Qy   2734 gatgctgcaggcgcgcaaaccggcagngtgtcagacacgccgcgccgccgttcg------ 2787
          ||||| ||||| ||||||||||||| | | ||||  ||||||||||||||||
Db   4483 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGT 4424

Qy   2788 ---cgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacg 2844
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4423 TCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 4364

Qy   2845 ctgacggtaaacggcaaattgaacngtcaaggaacattccgctttatgtcggaactcttc 2904
          |||||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||
Db   4363 CTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 4304

Qy   2905 ggctaccgaagcgacaaattgaagctggcggaaagttccgaaggnacttacaccttggcg 2964
          |||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   4303 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCG 4244

Qy   2965 gtcaacaataccggcaacgaacccgtaagcctcgatcaattgacggtagtggaagggaaa 3024
          |||||||||||||||||||||| | ||||||||| |||||||||||||||||||| |||
Db   4243 GTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 4184

Qy   3025 gacaacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgcc 3084
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4183 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCC 4124

Qy   3085 ggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtc 3144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4123 GGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC 4064

Qy   3145 aaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaa 3204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4063 AAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 4004

Qy   3205 gacaacgcgcaaagccttgacgcgctgattgcggccgggcgcgatgccgccgaaaagaca 3264
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   4003 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACA 3944

Qy   3265 gaaagcgttgccgaaccggcccggcnggcaggcggggaaaatgtcggcattatgcaggcg 3324
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   3943 GAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3884

Qy   3325 gaggaagagaaaaaacgggtgcaggcggataaagacagcgcnttggcgaaacagcgcgaa 3384
          |||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db   3883 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAA 3824

Qy   3385 gcggaaacccggccggntaccaccgccttccccccgcgcccgcngcgcccgccgggatttg 3444
          |||||||||||||||| ||||||||||||||||||||||| ||||||||||||||||||||
Db   3823 GCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTG 3764

Qy   3445 ccgcaaccgcagccccaaccgcaacctcaaccccaaccgcagcgcgacctgatnagccgt 3504
          ||||||| ||| |||||||||      || ||||||||||||||||||||||| ||||||
Db   3763 CCGCAACTGCAACCCCAACCG------CAGCCCCAACCGCAGCGCGACCTGATCAGCCGT 3710

Qy   3505 tatgccaatagcggtttgagtgaatttttccgccacgctcaacagcgttttcgccgtacag 3564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3709 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAG 3650

Qy   3565 gacgaattggaccgcgtgtttgccgaagaccgccgcaacgcngtttggacaagcngcatc 3624
          |||||||| |||||||| ||||||||||||||||||||||| |||||||||||| |||||
```

```
Db   3649 GACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATC 3590

Qy   3625 cggnacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctg 3684
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3589 CGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 3530

Qy   3685 cgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcac 3744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3529 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCAC 3470

Qy   3745 aaccggaccgaaaacancttcgacgacggcatcggcaactcggcacggcttgcccacggc 3804
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3469 AACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3410

Qy   3805 gccgttttcgggcaatacggcatcggcaggttcgacatcggcatcagcacgggcgcgggt 3864
          |||||||||||||||||||||||| ||||||| |||||||||||||| ||||||||||
Db   3409 GCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGT 3350

Qy   3865 tttagcagcggcantctntcagacggcatcggaggcaaaatccgccgccgcgtgctgcat 3924
          |||||||||||||  || ||||||||||||||||||||||||||||||||||||||||||
Db   3349 TTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCAT 3290

Qy   3925 tacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgtacatc 3984
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   3289 TACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATC 3230

Qy   3985 ggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgcc 4044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3229 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCC 3170

Qy   4045 acccccggtcttgcgttcaaccgntaccgngcgggcattaaggcagattattcattcaaa 4104
          |||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db   3169 ACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAA 3110

Qy   4105 ccggcgcaacacatntccatcacnccttatttnagcctgtcctataccgatgccgcttcg 4164
          ||||||||||||| |||||||| |||||||| |||||||||||||||||||||||||||||
Db   3109 CCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCG 3050

Qy   4165 ggcaaagtccgaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgc 4224
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   3049 GGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGC 2990

Qy   4225 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgcc 4284
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   2989 AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCC 2930

Qy   4285 gccgccaaaggnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4344
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2929 GCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGC 2870

Qy   4345 tggtaa 4350
          ||||||
Db   2869 TGGTAA 2864
```

```
RESULT   7
AX044034/c
LOCUS       AX044034               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 113 from Patent WO0066791.
ACCESSION   AX044034
VERSION     AX044034.1  GI:11342918
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 113 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
```

```
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    86473 a  95646 c  85908 g  81953 t
ORIGIN
```

Query Match 80.6%; Score 3508; DB 6; Length 349980;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 3876; Conservative 0; Mismatches 462; Indels 48; Gaps 8;

```
Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
Db 289191 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 289132

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
Db 289131 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 289072

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
Db 289071 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 289012

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
Db 289011 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 288952

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
Db 288951 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 288892

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
Db 288891 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 288832

Qy     361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
Db 288831 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 288772

Qy     421 gtgaaaagaaataattataagcctgacaattca---caccctacaacggcgattancat 477
Db 288771 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 288712

Qy     478 atgccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatg 537
Db 288711 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 288652

Qy     538 aggggggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcagga 597
Db 288651 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 288592

Qy     598 caccactattggcgttatgatgatga------caaacacggcgatttatcctactccggc 651
Db 288591 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 288532

Qy     652 gca------------tggttaattggcggcaatacacatatgcagggttggggaaataat 699
Db 288531 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 288472

Qy     700 ggcgtanttagtttgagcggcgatgtgcgccatgccaacgactatggccctatgccgatt 759
Db 288471 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 288412

Qy     760 gcaggtgcggcaggcgacagcggttcgccaatgtttatttatgacaaaacaaacaataaa 819
```

```
Db 288411 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 288352

Qy     820 tggctgctcaacggagttttacaaaccggctacccttattccggcagggaaaacggtttc 879
Db 288351 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 288292

Qy     880 cagctgatacgcaaagattggttctacgatgacatttacagaggcgatacacataccgtc 939
Db 288291 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 288232

Qy     940 tnttttgaaccgcgcagtaacggacatttttccttacatccaacaacaacggtacgggt 999
Db 288231 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGA 288172

Qy    1000 acggtaacagaaaccaacgaaaaggtntccaatccaaa---gcttaaagtacagacagtc 1056
Db 288171 AAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 288112

Qy    1057 cgactgtttgacgaatctttgaatgaaactgataaagaaccagttt---acgcggcaggg 1113
Db 288111 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 288052

Qy    1114 ggtgttaatcagtaccgtccaaggttaaacaacggtgaaaaccttttcttttatcgattac 1173
Db 288051 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 287992

Qy    1174 ggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcgggcggtttgtatttt 1233
Db 287991 GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 287932

Qy    1234 gaaggtgattttacggtctcgcctgaaaacaacgaaacgtggcaaggcgcgggcgttcat 1293
Db 287931 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 287872

Qy    1294 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1353
Db 287871 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 287812

Qy    1354 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1413
Db 287811 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 287752

Qy    1414 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1473
Db 287751 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 287692

Qy    1474 tttagtgaaatcggcttgntcagcggcaggggtacggtgcaactgaatgccgataatcag 1533
Db 287691 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 287632

Qy    1534 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1593
Db 287631 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 287572

Qy    1594 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgncnatcataat 1653
Db 287571 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 287512

Qy    1654 gccacaacaacatccaccgttaccattacagggaatgaaagtattacacaaccgagtggt 1713
Db 287511 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGG---- 287454

Qy    1714 aagaatatcaatagacttaattacagcaaagaaattgcctacaacggttggtttggcgag 1773
Db 287455 --CAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 287398

Qy    1774 aaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagac 1833
Db 287397 AAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGAC 287338

Qy    1834 cgcacccngctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggc 1893
Db 287337 CGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGC 287278
```



```
          |||||||| |||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db 285123 ACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAA 285064

Qy   4105 ccggcgcaacacatntccatcacncctttatttnagcctgtcctataccgatgccgcttcg 4164
          ||||||||||||||| |||||||| |||||||| ||||||||||||||||||||||||||||
Db 285063 CCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCG 285004

Qy   4165 ggcaaagtccgaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgc 4224
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 285003 GGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGC 284944

Qy   4225 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgcc 4284
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 284943 AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCC 284884

Qy   4285 gccgccaaaggnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4344
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db 284883 GCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGC 284824

Qy   4345 tggtaa 4350
          ||||||
Db 284823 TGGTAA 284818
```

```
RESULT   8
NME242535
LOCUS       NME242535               4374 bp    DNA     linear   BCT 16-JAN-2002
DEFINITION  Neisseria meningitidis serogroup B app gene, strain B:15:P1.16.
ACCESSION   AJ242535
VERSION     AJ242535.2  GI:6977940
KEYWORDS    APP gene; app protein.
SOURCE      Neisseria meningitidis serogroup B.
  ORGANISM  Neisseria meningitidis serogroup B
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (sites)
  AUTHORS   Ala Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1999) Ala Aldeen D.A., Microbiology, University
            of Nottingham, Division of Microbiology, SCLS, University Hospital,
            Nottingham, NG7 2UH, UNITED KINGDOM
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 4374)
  AUTHORS   Ala Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) Ala Aldeen D.A., Microbiology, University
            of Nottingham, Division of Microbiology, SCLS, University Hospital,
            Nottingham, NG7 2UH, UNITED KINGDOM
COMMENT     On Feb 14, 2000 this sequence version replaced gi:4938471.
FEATURES             Location/Qualifiers
     source          1. .4374
                     /organism="Neisseria meningitidis serogroup B"
                     /strain="B:15:P1.16"
                     /isolate="SD"
                     /db_xref="taxon:491"
     gene            1. .4374
                     /gene="app"
     CDS             1. .4374
                     /gene="app"
                     /codon_start=1
                     /transl_table=11
                     /product="App protein"
                     /protein_id="CAB43832.2"
                     /db_xref="GI:6977941"
                     /db_xref="SPTREMBL:Q9X7H1"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDAHSVFYEPRQNGKYSFND
                     DNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINQGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGTVILDQQADDKGKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGSLSANGDTRYTVSHYATQNG
                     NLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHNAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFNTLTVNGKLNGQGTFR
                     FMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFT
                     LQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALI
                     AAGRDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATT
                     AFPRARRARRDLPQLQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFA
                     EDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENT
                     FDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVLHYGIQ
                     ARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA
                     QHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAA
                     AAKGPQLEAQHSAGIKLGYRW"
BASE COUNT     1265 a    1163 c    1066 g     880 t
ORIGIN
```

```
  Query Match             80.5%;  Score 3500;  DB 1;  Length 4374;
  Best Local Similarity   88.3%;  Pred. No. 0;
  Matches 3871;  Conservative    0;  Mismatches  467;  Indels   48;  Gaps    8;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy     361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
           |||||||||||||||||||||||||||||| || |||||| ||||  |||||
Db     361 AACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy     421 gtgaaaagaaataattataagcctgacaattca---caccctacaacggcgattancat 477
           |||||| | |||||||||||  | |  | |  |   || |||||   ||||||||| |||
Db     421 GTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCAT 480

Qy     478 atgccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatg 537
           |||||||||||||||||||||||||||||||||||||||| |||||||| ||| | |||
Db     481 ATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATG 540

Qy     538 agggggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcagga 597
              |||   |  |||  |||| || | || || || || || ||||| || || ||  ||||
Db     541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600

Qy     598 caccactattggcgttatgatgatga------caaacacggcgatttatcctactccggc 651
```



```
Qy   2785 tcgcgccgttccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacg 2844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2815 TCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACG 2874

Qy   2845 ctgacggtaaacggcaaattgaacngtcaaggaacattccgctttatgtcggaactcttc 2904
          |||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db   2875 CTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934

Qy   2905 ggctaccgaagcgacaaattgaagctggcggaaagttccgaaggnacttacaccttggcg 2964
          |||||||| |||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCG 2994

Qy   2965 gtcaacaataccggcaacgaacccgtaagcctcgatcaattgacggtagtggaagggaaa 3024
          |||||||||||||||||||| | |||||||| |||||||||||||||||||| |||
Db   2995 GTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054

Qy   3025 gacaacaaaccgctgtccgaaaacctttaatttcaccctgcaaaacgaacacgtcgatgcc 3084
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3055 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCC 3114

Qy   3085 ggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtc 3144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3115 GGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC 3174

Qy   3145 aaagaacaagagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaa 3204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3175 AAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 3234

Qy   3205 gacaacgcgcaaaqccttgacgcgctgattgcggccgggcgcgatgccgccgaaaagaca 3264
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   3235 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACA 3294

Qy   3265 gaaagcgttgccgaaccggcccggcnggcaggcggggaaaatgtcggcattatgcaggcg 3324
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3295 GAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3354

Qy   3325 gaggaagagaaaaaacgggtgcaggcggataaagacagcgcnttggcgaaacagcgcgaa 3384
          |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db   3355 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAA 3414

Qy   3385 gcggaaacccggccggntaccaccgccttccccgcgcccgcngcgcccgccgggatttg 3444
          |||||||||||||||| ||||||||||||||||||||||| ||||||||||||||||||||
Db   3415 GCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTG 3474

Qy   3445 ccgcaaccgcagccccaaccgcaacctcaaccccaaccgcagcgcgacctgatnagccgt 3504
          ||||||| ||| |||||||||      || ||||||||||||||||||||| ||||||
Db   3475 CCGCAACTGCAACCCCAACCG------CAGCCCCAACCGCAGCGCGACCTGATCAGCCGT 3528

Qy   3505 tatgccaatagcggtttgagtgaatttccgccacgctcaacagcgttttcgccgtacag 3564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3529 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAG 3588

Qy   3565 gacgaattggaccgcgtgtttgccgaagaccgccgcaacgcngtttggacaagcngcatc 3624
          |||||||| |||||||| |||||||||||||||||||||| |||||||||||| |||||
Db   3589 GACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATC 3648

Qy   3625 cggnacaccaaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctg 3684
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3649 CGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 3708

Qy   3685 cgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcac 3744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3709 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCAC 3768

Qy   3745 aaccggaccgaaaacancttcgacgacggcatcggcaactcggcacggcttgcccacggc 3804
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3769 AACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3828

Qy   3805 gccgttttcgggcaatacggcatcggcaggttcgacatcggcatcagcacgggcgcgggt 3864
          |||||||||||||||||||||||| ||||||| |||||||||||||| ||||||||||||
Db   3829 GCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGT 3888

Qy   3865 tttagcagcggcantctntcagacggcatcggaggcaaaatccgccgccgcgtgctgcat 3924
          |||||||||||||  || ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3889 TTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCAT 3948

Qy   3925 tacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgtacatc 3984
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3949 TACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATC 4008

Qy   3985 ggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgcc 4044
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4009 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCC 4068

Qy   4045 acccccggtcttgcgttcaaccgntaccgngcgggcattaaggcagattattcattcaaa 4104
          |||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||||||
Db   4069 ACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAA 4128

Qy   4105 ccggcgcaacacatntccatcacncctttatttnagcctgtcctataccgatgccgcttcg 4164
          |||||||||||||| |||||||| |||||||| ||||||||||||||||||||||||||||||||
Db   4129 CCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCG 4188

Qy   4165 ggcaaagtccgaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgc 4224
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   4189 GGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGC 4248

Qy   4225 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgcc 4284
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   4249 AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCC 4308

Qy   4285 gccgccaaaggnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4344
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4309 GCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGC 4368

Qy   4345 tggtaa 4350
          ||||||
Db   4369 TGGTAA 4374
```

```
RESULT   9
NME296277
LOCUS       NME296277               4386 bp    DNA     linear   BCT 14-JAN-2002
DEFINITION  Neisseria meningitidis app gene for adhesion penetration protein,
            strain Z4181.
ACCESSION   AJ296277
VERSION     AJ296277.1  GI:11071864
KEYWORDS    adhesion penetration protein; APP gene.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (sites)
  AUTHORS   Hadi,H.A., Wooldridge,K.G., Robinson,K. and Ala'Aldeen,D.A.
  TITLE     Identification and characterization of App: an immunogenic
            autotransporter protein of Neisseria meningitidis
  JOURNAL   Mol. Microbiol. 41 (3), 611-623 (2001)
  MEDLINE   21424673
REFERENCE   2  (bases 1 to 4386)
  AUTHORS   Ala-Aldeen,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2000) Ala-Aldeen D.A., Microbiology, University
            of Nottingham, Molecular Bacteriology and Immunology Group,
            University Hospital, Nottingham NG7 2UH, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .4386
                     /organism="Neisseria meningitidis"
                     /strain="Z4181"
                     /db_xref="taxon:487"
                     /country="Mali"
                     /note="serogroup C"
     gene            1. .4365
                     /gene="app"
     CDS             1. .4365
                     /gene="app"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="adhesion penetration protein"
                     /protein_id="CAC14671.1"
                     /db_xref="GI:11071865"
                     /db_xref="SPTREMBL:Q9F3Z4"
                     /translation="MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGH
                     TYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVA
                     ALVGDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFTYKIVKRNNYKAGTNGHPYGGDYH
                     MPRLHKFVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSY
                     HIASAYSWLVGGNTFAQNGSGGGTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDA
                     QKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVFYEPRQNGKYSFHD
                     NNNGTGKIDAKHEHNSLPNRLKTRTVQLFNVSLSETAQEPVYHAAGGVNSYRPRLNNG
                     ENISFIDEGKGELILTSNINRGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWK
                     VNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGKVILDQQADENNKKQAFSEIGLVS
                     GRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIVNHNQDKEST
                     VTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLS
                     GGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKA
                     ENFQIKGGQAVVSRNVAKVKGDWHLSNHAQAVFGVAPHQSHTICTRSDWTGLTNCVEK
                     TITDDKVIASLTNTDISGNVSLADHAHLNLTGLATLNGNLSANGDTRYTVSHNATQNG
                     DLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNG
                     NVSLADKAVFHSESSRFTGQISGSKDTALHLKDSEWTLPSGTELGNLNLDNATITLNS
                     AYRHDAAGAQTGSATDAPRRRSRRSLLSVTPPASVESRFNTLTVNGKLNGQGTFRFMS
                     ELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKPLSENLNFTLQN
                     EHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALIAAG
                     RDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATTAFP
                     RARRARRDLPQLQPQPQPQPQPQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEGR
                     RNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDD
                     GIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGSKIRRRVLHYGIQARY
                     RAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
                     SITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAK
                     GPQLEAQHSAGIKLGYRW"
BASE COUNT     1270 a    1167 c    1066 g     883 t
ORIGIN

  Query Match             80.2%;  Score 3490.2;  DB 1;  Length 4386;
  Best Local Similarity   88.0%;  Pred. No. 0;
  Matches 3853;  Conservative    0;  Mismatches  485;  Indels   39;  Gaps    7;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGC 60

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCT 120

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy    361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaatt 420
          |||||||||||||||||||||||||| |||||||||| || |||||| |||| |||||
Db    361 AACGTTGATTTTGGTGCGGAAGGAAGCAATCCCGATCAACATCGTTTTACTTATAAAATT 420

Qy    421 gtgaaaagaaataattataagcctg---acaattcacacccttacaacggcgattancat 477
          |||||| | |||||||||||  | |     ||    || |||||   |||||||| |||
Db    421 GTGAAACGGAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCAT 480

Qy    478 atgccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatg 537
          |||||||||||| |||||||||||| || |||||||||||||| |||||||| ||| | |||
Db    481 ATGCCGCGTTTACATAAATTTGTAACCGATGCAGAACCTGTTGAAATGACCAGTTATATG 540

Qy    538 aggggggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcagga 597
            |||   |  |||  |||| || | || || || || ||||| || || ||  ||||
Db    541 GATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600

Qy    598 caccactattggcgttatgatgatga------caaacacggcgatttatcctactccggc 651
            || |||||||| | |||||| ||      |||  || ||||    || ||
Db    601 AGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATT 660

Qy    652 gca------------tggttaattggcggcaatacacatatgcagggttggggaaataat 699
          |||             ||| |  |||| ||||||||   |   ||   | |   | |  |
Db    661 GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy    700 ggcgtanttagtttgagcggcgatgtgcgccatgccaacgactatggccctatgccgatt 759
          |||  | | |  ||  |  | ||       |   | |  |||||  | | || |
Db    721 GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780

Qy    760 gcaggtgcggcaggcgacagcggttcgccaatgtttatttatgacaaaacaaacaataaa 819
          | |||  |    |||||||| || || ||||||||||| |||||      |||  | |||
Db    781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAA 840

Qy    820 tggctgctcaacggagttttacaaaccggctacccttattccggcagggaaaacggtttc 879
          |||  |  | || || || || ||||| ||| |||| |||   || |     || || |||
Db    841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

Qy    880 cagctgatacgcaaagattggttctacgatgacatttacagaggcgatacacataccgtc 939
          |||||| | || ||||||||||||||| ||||| || |     || |||| ||| | ||
Db    901 CAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960

Qy    940 tnttttgaaccgcgcagtaacggacatttttccttacatccaacaacaacggtacgggt 999
          | |  ||||| ||    || || | |  || |||     ||| || || || || || ||
Db    961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTCACGACAATAATAATGGCACAGGA 1020

Qy   1000 acggtaacagaaaccaacgaaaaggtntccaatccaaa---gcttaaagtacagacagtc 1056
          |  |    |  |   | ||| |    ||    || ||     | |||  ||  || ||
Db   1021 AAAATCGACGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080

Qy   1057 cgactgtttgacgaatctttgaatgaaactgataaagaaccagttt---acgcggcaggg 1113
          | | ||||| | |  |||||    || || |   |||||| ||||   | || |||||
Db   1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCACAAGAACCTGTTTATCATGCTGCAGGT 1140

Qy   1114 ggtgttaatcagtaccgtccaaggttaaacaacggtgaaaacctttctttttatcgattac 1173
          |||||  ||    || || || ||  | || || || |||||  |||| ||||| ||  |
Db   1141 GGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAA 1200

Qy   1174 ggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcgggcggtttgtatttt 1233
          || || ||| || | ||  |  | |  ||||||||| ||| || || || || |||||
Db   1201 GGAAAAGGCGAATTGATACTTACCAGTAACATCAACCGAGGTGCTGGAGGATTATATTTC 1260

Qy   1234 gaaggtgattttacggtctcgcctgaaaacaacgaaacgtggcaaggcgcgggcgttcat 1293
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CAAGGAGATTTTACGGTCTCGCCTGAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320

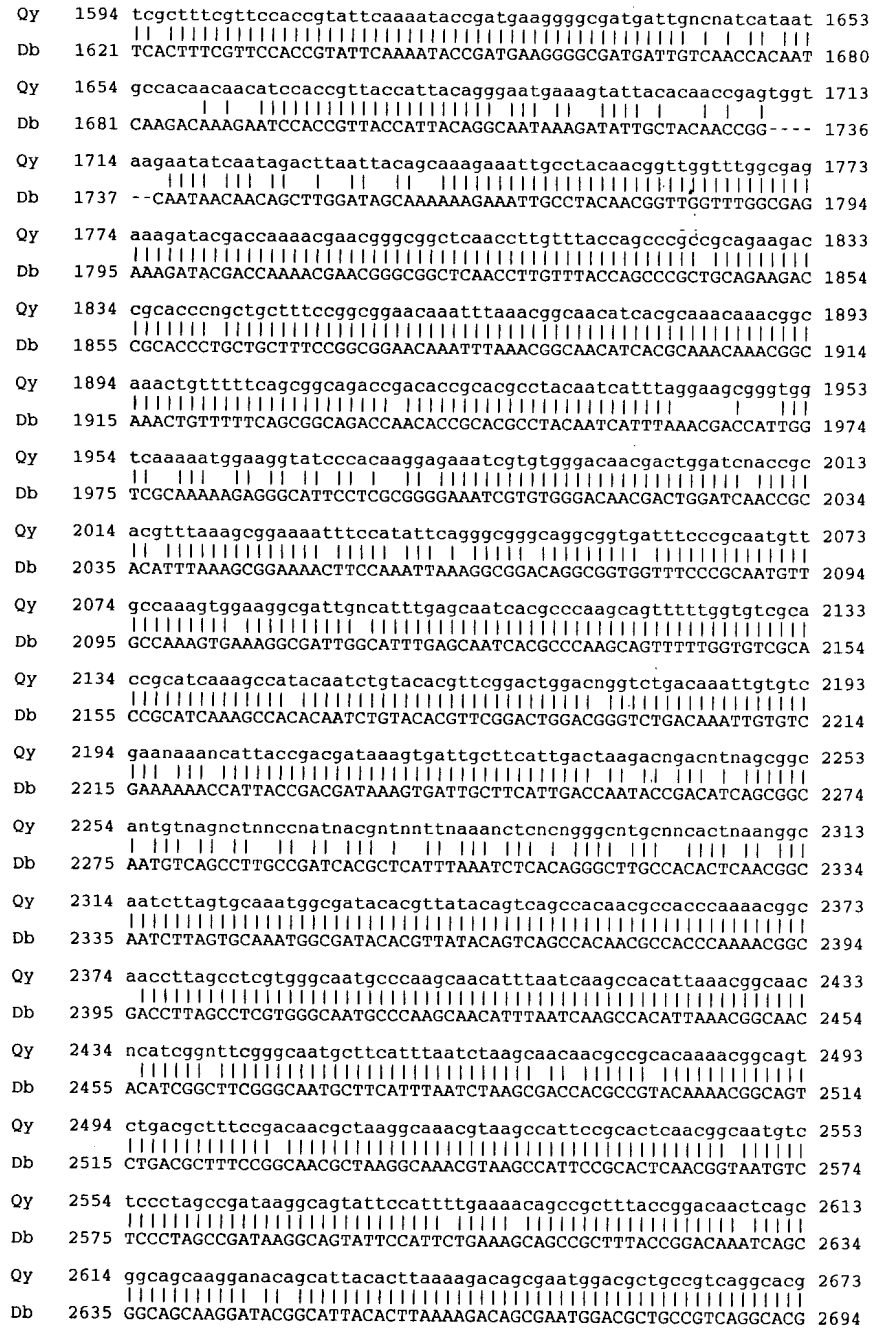
Qy   1294 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1353
          |||||||||||||||||||||||||||||||||| || ||||||||||||||||||||||
Db   1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTCAATGGCGTGGCAAACGACCGCCTGTCC 1380

Qy   1354 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440

Qy   1414 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1473
          |||||||||||| ||| |||| |||| || |||| || ||    ||||||||||||
Db   1441 GTGGGCGACGGTAAAGTTATTTTAGATCAACAAGCAGATGAAAATAATAAAAAACAAGCC 1500

Qy   1474 tttagtgaaatcggcttgntcagcggcaggggtacggtgcaactgaatgccgataatcag 1533
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560

Qy   1534 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1593
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
```


Qy 1594 tcgctttcgttccaccgtattcaaaatacCgatgaaggggcgatgattgtcnatcataat 1653
Db 1621 TCACTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
Qy 1654 gccacaacaacatccaccgttaccattacagggaatgaaagtattacacaaccgagtggt 1713
Db 1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGG---- 1736
Qy 1714 aagaatatcaatagacttaattacagcaaagaaattgcctacaacggttggtttggcgag 1773
Db 1737 --CAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1794
Qy 1774 aaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagac 1833
Db 1795 AAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCTGCAGAAGAC 1854
Qy 1834 cgcacccngctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggc 1893
Db 1855 CGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGC 1914
Qy 1894 aaactgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtgg 1953
Db 1915 AAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGG 1974
Qy 1954 tcaaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgactggatcnaccgc 2013
Db 1975 TCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGC 2034
Qy 2014 acgtttaaagcggaaaatttccatattcagggcgggcaggcggtgatttcccgcaatgtt 2073
Db 2035 ACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTT 2094
Qy 2074 gccaaagtggaaggcgattgncatttgagcaatcacgcccaagcagtttttggtgtcgca 2133
Db 2095 GCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCA 2154
Qy 2134 ccgcatcaaagccatacaatctgtacacgttcggactggacnggtctgacaaattgtgtc 2193
Db 2155 CCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 2214
Qy 2194 gaanaaancattaccgacgataaagtgattgcttcattgactaagacngacntnagcggc 2253
Db 2215 GAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACCAATACCGACATCAGCGGC 2274
Qy 2254 antgtnagnctnnccnatnacgntnnttnaaanctcncngggcntgcnncactnaanggc 2313
Db 2275 AATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 2334
Qy 2314 aatcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggc 2373
Db 2335 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGC 2394
Qy 2374 aaccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2433
Db 2395 GACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 2454
Qy 2434 ncatcggnttcgggcaatgcttcatttaatctaagcaacaacgccgcacaaaacggcagt 2493
Db 2455 ACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGT 2514
Qy 2494 ctgacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtc 2553
Db 2515 CTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTC 2574
Qy 2554 tccctagccgataaggcagtattccattttgaaaacagccgctttaccggacaactcagc 2613
Db 2575 TCCCTAGCCGATAAGGCAGTATTCCATTCTGAAAGCAGCCGCTTTACCGGACAAATCAGC 2634
Qy 2614 ggcagcaaggaNacagcattacacttaaaagacagcgaatggacgctgccgtcaggcacg 2673
Db 2635 GGCAGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACG 2694
Qy 2674 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccac 2733
Db 2695 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2754
Qy 2734 gatgctgcaggcgcgcaaaccggcagngtgtcagacacgccgcgccgccgttcgcgccgt 2793
Db 2755 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814
Qy 2794 tccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacggta 2853
Db 2815 TCCCTATTATCCGTTACACCTCCGGCTTCGGTAGAATCCCGTTTCAACACGCTGACGGTA 2874
Qy 2854 aacggcaaattgaacngtcaaggaacattccgctttatgtcggaactcttcggctaccga 2913
Db 2875 AACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGC 2934
Qy 2914 agcgacaaattgaagctggcggaaagttccgaaggnacttacaccttggcggtcaacaat 2973
Db 2935 AGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT 2994
Qy 2974 accggcaacgaacccgtaagcctcgatcaattgacggtagtggaagggaaagacaacaaa 3033
Db 2995 ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAA 3054
Qy 3034 ccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcgtgg 3093
Db 3055 CCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGG 3114
Qy 3094 cgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaa 3153
Db 3115 CGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA 3174
Qy 3154 gagctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaagacaacgcg 3213
Db 3175 GAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCG 3234
Qy 3214 caaagccttgacgcgctgattgcggccgggcgcgatgccgccgaaaagacagaaagcgtt 3273
Db 3235 CAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTT 3294
Qy 3274 gccgaaccggcccggcnggcaggcgggggaaaatgtcggcattatgcaggcggaggaagag 3333
Db 3295 GCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAG 3354
Qy 3334 aaaaaacgggtgcaggcggataaagacagcgcnttggcgaaacagcgcgaagcggaaacc 3393
Db 3355 AAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACC 3414
Qy 3394 cggccggntaccaccgccttcccccgcgcccgcngcgcccgccgggatttgccgcaaccg 3453
Db 3415 CGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTG 3474
Qy 3454 cagccccaaccgcaacctcaaccccaaccgcagcgcgacctgatnagccgttatgccaat 3513
Db 3475 CAACCCCAACCG------CAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAAT 3528
Qy 3514 agcggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacaggacgaattg 3573
Db 3529 AGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTA 3588
Qy 3574 gaccgcgtgtttgccgaagaccgccgcaacgcngtttggacaagcngcatccggnacacc 3633
Db 3589 GACCGCGTATTTGCCGAAGGCCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACC 3648
Qy 3634 aaacactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatc 3693
Db 3649 AAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATC 3708
Qy 3694 ggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggacc 3753
Db 3709 GGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACC 3768
Qy 3754 gaaaacancttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttc 3813

```
Db   3769 GAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTC 3828

Qy   3814 gggcaatacggcatcggcaggttcgacatcggcatcagcacgggcgcgggttttagcagc 3873
          |||||||||||||||| ||||||| |||||||||||||| ||||||||||||||||||||
Db   3829 GGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGC 3888

Qy   3874 ggcantctntcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcatt 3933
          |||| ||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   3889 GGCAGTCTTTCAGACGGCATCGGAAGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATT 3948

Qy   3934 caggcacgataccgcgccggtttcggcggattcggcatcgaaccgtacatcggcgcaacg 3993
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   3949 CAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACG 4008

Qy   3994 cgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggt 4053
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4009 CGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGC 4068

Qy   4054 cttgcgttcaaccgntaccgngcgggcattaaggcagattattcattcaaaccggcgcaa 4113
          |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db   4069 CTTGCGTTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAA 4128

Qy   4114 cacatntccatcacncctttatttnagcctgtcctataccgatgccgcttcgggcaaagtc 4173
          ||||| |||||||| |||||||| ||||||||||||||||||||||||||||||||||||
Db   4129 CACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTC 4188

Qy   4174 cgaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgcagtgcggaa 4233
          ||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   4189 CGAACGCGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAA 4248

Qy   4234 tggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgccgccgccaaa 4293
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   4249 TGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAA 4308

Qy   4294 ggnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4350
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4309 GGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4365


RESULT  10
A96620
LOCUS       A96620                  4407 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 653 from Patent WO9924578.
ACCESSION   A96620
VERSION     A96620.1  GI:6780202
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4407)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 653 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4407
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     1273 a   1169 c   1094 g    871 t
ORIGIN


  Query Match           77.8%;  Score 3385.8;  DB 6;  Length 4407;
  Best Local Similarity  86.3%;  Pred. No. 0;
  Matches 3814;  Conservative    0;  Mismatches  524;  Indels   81;  Gaps    7;

Qy      1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAAAACCGGCCGC 60

Qy     61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
          |||||||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db     61 ATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCCCCAAGCC 120

Qy    121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy    181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy    241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
          |||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||||||
Db    241 GTCGGCAAATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300

Qy    301 gtggcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataac 360
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAAC 360

Qy    361 aacgttgattttggtgcggaaggaagnaatcccgatcagcaccgttttttcttaccaaatt 420
          || |||||||||||||||| ||||| |||||||||||||||||| ||||||||||||||
Db    361 AATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATT 420

Qy    421 gtgaaaagaaataattataagcctg---acaattcacacccttacaacggcgattancat 477
          |||||||||||||||||| | |      ||    || |||||   |||||||| |||
Db    421 GTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCAT 480

Qy    478 atgccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatg 537
          |||||||||||||| |||||||||||||||||||||||||||||| || ||||| ||| |||
Db    481 ATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATG 540

Qy    538 agggggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcagga 597
            |||   |  ||  | |||  | | ||||| || || ||||| || |||||  ||||
Db    541 GATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGC 600

Qy    598 caccactattggcgttatgatgatgacaaacacggcg----------------attta 639
             || |||||||| | |||||| ||| ||| |                    || |
Db    601 AGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATT 660

Qy    640 tcctactccggcgcatggttaattggcggcaatacacatatgcagggttggggaaataat 699
          |  | |    | ||| |  | || |||||||||   |   ||  | |   |  |  |
Db    661 GCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT 720

Qy    700 ggcgtanttagtttgagcggcgatgtgcgccatgccaacgactatggccctatgccgatt 759
          |||  | | |  ||  | ||||        |    | |   |||||   | | || |
Db    721 GGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780

Qy    760 gcaggtgcggcaggcgacagcggttcgccaatgtttatttatgacaaaacaaacaataaa 819
          | ||| |    |||||||| || || ||||||||||| |||||      |||  | ||
Db    781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840

Qy    820 tggctgctcaacggagttttacaaaccggctacccttattccggcagggaaaacggtttc 879
          ||| |  | || || || || ||||| ||| |||| |||   || |     || || |||
Db    841 TGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900

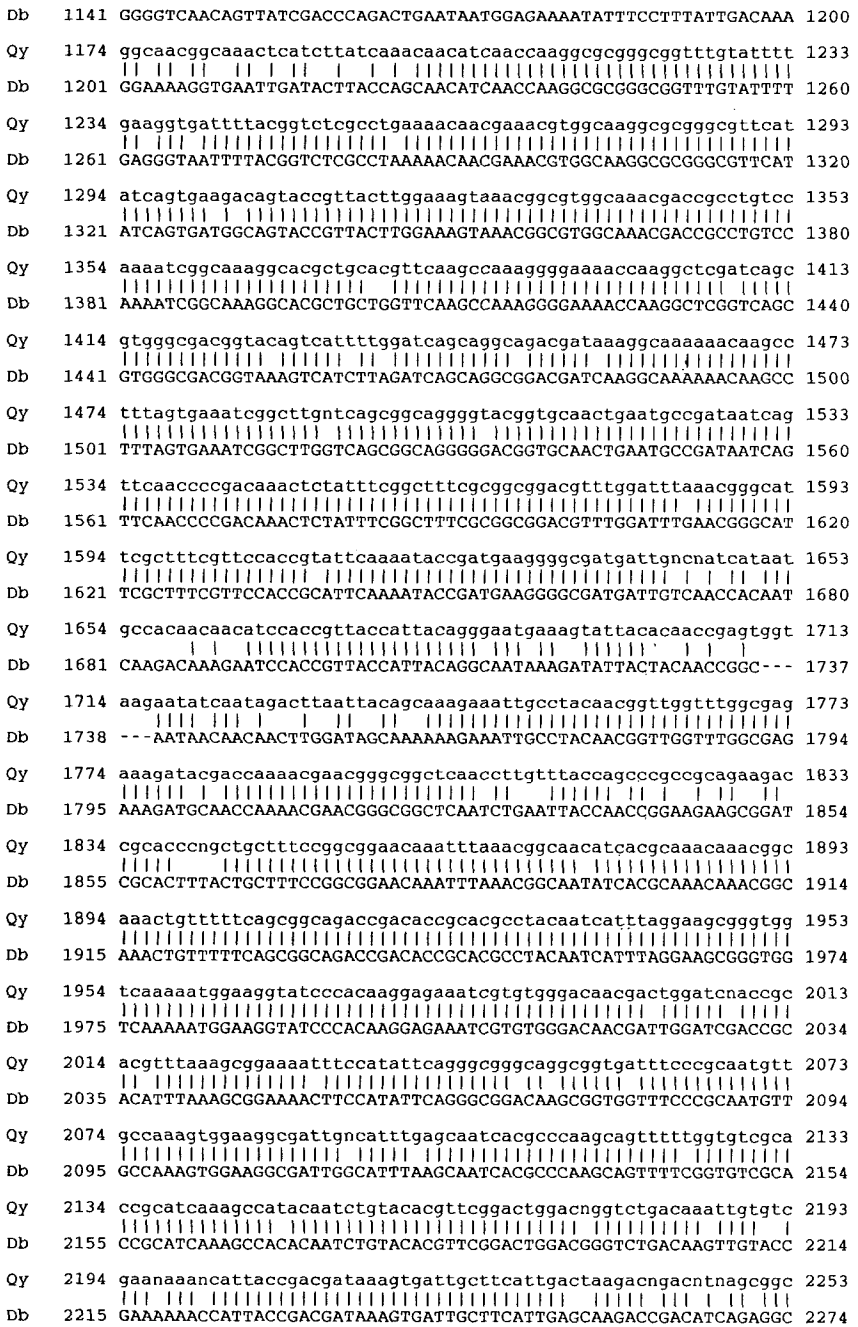
Qy    880 cagctgatacgcaaagattggttctacgatgacatttacagaggcgatacacatatccgtc 939
          |||||  | || |||||||||||||| ||||| || |     || ||||| ||| | ||
Db    901 CAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTA 960

Qy    940 tnttttgaaccgcgcagtaacggacatttttccttttacatccaacaacaacggtacgggt 999
          |  |  ||||| |     || ||  | |  |  ||||    ||| || || ||  | ||
Db    961 TTCTACGAACCACATCAAAATGGGAAATACTTTTTTTAACGACAATAATAATGGCGCAGGA 1020

Qy   1000 acggtaacagaaaccaacgaaaaggtntccaatcc---aaagcttaaagtacagacagtc 1056
          |   |    |  |   || |    ||    ||    |   | ||| ||  || ||
Db   1021 AAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTT 1080

Qy   1057 cgactgtttgacgaatctttgaatgaaactgataaagaaccagttt---acgcggcaggg 1113
          | | ||||| | |  |||||    || || |  | |||||| ||||    | || |||||
Db   1081 CAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140

Qy   1114 ggtgttaatcagtaccgtccaaggttaaacaacggtgaaaaccttttcttttatcgattac 1173
          || || ||    || || || ||  | || || || |||||  |||| ||||| ||  |
```


Db 1141 GGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAA 1200
Qy 1174 ggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcgggcggtttgtatttt 1233
Db 1201 GGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 1260
Qy 1234 gaaggtgattttacggtctcgcctgaaaacaacgaaacgtggcaaggcgcgggcgttcat 1293
Db 1261 GAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
Qy 1294 atcagtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtcc 1353
Db 1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCC 1380
Qy 1354 aaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagc 1413
Db 1381 AAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGC 1440
Qy 1414 gtgggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagcc 1473
Db 1441 GTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCC 1500
Qy 1474 tttagtgaaatcggcttgntcagcggcaggggtacggtgcaactgaatgccgataatcag 1533
Db 1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAG 1560
Qy 1534 ttcaaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcat 1593
Db 1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
Qy 1594 tcgctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgncnatcataat 1653
Db 1621 TCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
Qy 1654 gccacaacaacatccaccgttaccattacagggaatgaaagtattacacaaccgagtggt 1713
Db 1681 CAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGC--- 1737
Qy 1714 aagaatatcaatagacttaattacagcaaagaaattgcctacaacggttggtttggcgag 1773
Db 1738 ---AATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1794
Qy 1774 aaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagac 1833
Db 1795 AAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGAT 1854
Qy 1834 cgcacccngctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggc 1893
Db 1855 CGCACTTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGGC 1914
Qy 1894 aaactgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtgg 1953
Db 1915 AAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGG 1974
Qy 1954 tcaaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgactggatcnaccgc 2013
Db 1975 TCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGC 2034
Qy 2014 acgtttaaagcggaaaatttccatattcagggcgggcaggcggtgatttcccgcaatgtt 2073
Db 2035 ACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTT 2094
Qy 2074 gccaaagtggaaggcgattgncatttgagcaatcacgcccaagcagtttttggtgtcgca 2133
Db 2095 GCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCA 2154
Qy 2134 ccgcatcaaagccatacaatctgtacacgttcggactggacnggtctgacaaattgtgtc 2193
Db 2155 CCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACC 2214
Qy 2194 gaanaaancattaccgacgataaagtgattgcttcattgactaagacngacntnagcggc 2253
Db 2215 GAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGC 2274
Qy 2254 antgtnagnctnnccnatnacgntnnttnaaanctcncngggcntgcnncactnaanggc 2313
Db 2275 AATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGC 2334
Qy 2314 aatcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggc 2373
Db 2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGC 2394
Qy 2374 aaccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2433
Db 2395 AACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC 2454
Qy 2434 ncatcggnttcgggcaatgcttcatttaatctaagcaacaacgccgcacaaaacggcagt 2493
Db 2455 ACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGT 2514
Qy 2494 ctgacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtc 2553
Db 2515 CTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTC 2574
Qy 2554 tccctagccgataaggcagtattccattttgaaaacagccgctttaccggacaactcagc 2613
Db 2575 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGC 2634
Qy 2614 ggcagcaaggancagcattacacttaaaagacagcgaatggacgctgccgtcaggcacg 2673
Db 2635 GGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACG 2694
Qy 2674 gaattaggcaatttaaaccttgacaacgccaccattacactcaattccgcctatcgccac 2733
Db 2695 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
Qy 2734 gatgctgcaggcgcgcaaaccggcagngtgtcagacacgccgcgccgccgttcgcgccgt 2793
Db 2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814
Qy 2794 tccctattatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacggta 2853
Db 2815 TCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTA 2874
Qy 2854 aacggcaaattgaacngtcaaggaacattccgctttatgtcggaactcttcggctaccga 2913
Db 2875 AACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGC 2934
Qy 2914 agcgacaaattgaagctggcggaaagttccgaaggnacttacaccttggcggtcaacaat 2973
Db 2935 AGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAAT 2994
Qy 2974 accggcaacgaacccgtaagcctcgatcaattgacggtagtggaagggaaagacaacaaa 3033
Db 2995 ACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACA 3054
Qy 3034 ccgctgtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcgtgg 3093
Db 3055 CCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGG 3114
Qy 3094 cgttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaa 3153
Db 3115 CGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA 3174
Qy 3154 gagctttccgacaaactcggcaaggcagaa---------- 3183
Db 3175 GAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAG 3234
Qy 3184 ----------gccaaaaaacaggcggaaaaagacaacgcgcaaagccttgacgcgctg 3231
Db 3235 GCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTG 3294
Qy 3232 attgcggccgggcgcgatgccgccgaaaagacagaaagcgttgccgaaccggcccggcng 3291
Db 3295 ATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAG 3354

```
Qy   3292 gcaggcggggaaaatgtcggcattatgcaggcggaggaagagaaaaaacgggtgcaggcg 3351
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3355 GCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCG 3414

Qy   3352 gataaagacagcgcnttggcgaaacagcgcgaagcggaaacccggccggntaccaccgcc 3411
          |||||||||| ||| ||||||||||||||||||||||||||||||||| ||||||||||
Db   3415 GATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCC 3474

Qy   3412 ttcccccgcgcccgcngcgcccgccgggatttgccgcaaccgcagccccaaccgcaacct 3471
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3475 TTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCG------ 3528

Qy   3472 caaccccaaccgcagcgcgacctgatnagccgttatgccaatagcggtttgagtgaattt 3531
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   3529 CAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTT 3588

Qy   3532 tccgccacgctcaacagcgttttcgccgtacaggacgaattggaccgcgtgtttgccgaa 3591
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3589 TCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAA 3648

Qy   3592 gaccgccgcaacgcngtttggacaagcngcatccggnacaccaaacactaccgttcgcaa 3651
          |||||||||||||| |||||||||||| |||||||| ||||||||||||||||||||||
Db   3649 GACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAA 3708

Qy   3652 gatttccgcgcctaccgccaacaaaccgacctgcgccaaatcggtatgcagaaaaacctc 3711
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3709 GATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC 3768

Qy   3712 ggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccgaaaacancttcgacgac 3771
          |||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db   3769 GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGAC 3828

Qy   3772 ggcatcggcaactcggcacggcttgcccacggcgccgttttcgggcaatacggcatcggc 3831
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   3829 GGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGC 3888

Qy   3832 aggttcgacatcggcatcagcacgggcgcgggttttagcagcggcantctntcagacggc 3891
          |||||||||||||||||||| |||||||||||||||| ||||||| || |||||||||||
Db   3889 AGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGC 3948

Qy   3892 atcggaggcaaaatccgccgccgcgtgctgcattacggcattcaggcacgataccgcgcc 3951
          ||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   3949 ATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCA 4008

Qy   3952 ggtttcggcggattcggcatcgaaccgtacatcggcgcaacgcgctatttcgtccaaaaa 4011
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   4009 GGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAA 4068

Qy   4012 gcggattaccgctacgaaaacgtcaatatcgccacccccggtcttgcgttcaaccgntac 4071
          |||||||||||| |||||||||||||||||||||||| || ||||| |||||||| |||
Db   4069 GCGGATTACCGATACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTAC 4128

Qy   4072 cgngcgggcattaaggcagattattcattcaaaccggcgcaacacatntccatcacncct 4131
          || |||||||||||||||||||||||||||||||||||||||||||||| |||||||| |||
Db   4129 CGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCT 4188

Qy   4132 tatttnagcctgtcctataccgatgccgcttcgggcaaagtccgaacacgcgtcaatacc 4191
          ||||| |||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db   4189 TATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACC 4248

Qy   4192 gcngtattggctcaggatttcggcaaaacccgcagtgcggaatggggcgtaaacgccgaa 4251
          || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4249 GCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAA 4308

Qy   4252 atcaaaggtttcacgctgtccntccacgctgccgccgccaaaggnccgcaactggaagcg 4311
          |||||||||||||||||||| |||||||||||||||||| || |||||| ||||||||||
Db   4309 ATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCG 4368

Qy   4312 caacacagcgcgggcatcaaattaggctaccgctggtaa 4350
          || ||||||||||||||||||||||||||||||||||
Db   4369 CAGCACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4407
```

```
RESULT  11
A96614
LOCUS       A96614                  4355 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 647 from Patent WO9924578.
ACCESSION   A96614
VERSION     A96614.1  GI:6780199
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4355)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 647 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .4355
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      840 a    775 c    694 g    616 t   1430 others
ORIGIN


  Query Match           49.1%;  Score 2135.4;  DB 6;  Length 4355;
  Best Local Similarity  58.4%;  Pred. No. 0;
  Matches 2590;  Conservative   11;  Mismatches 1668;  Indels  167;  Gaps   11;

Qy       1 atgaaaacaaccgacaaacggacaaccgaaacacaccgcaaagccccgaaaaccggccgc 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGTCGC 60

Qy      61 atccgcttctcgcctgcttacttagccatatgcctgtcgttcggcattcttccccaagct 120
           |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCCGCTTCTCNGCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCC 120

Qy     121 tgggcgggacacacttatttcggcatcaactaccaatactatcgcgactttgccgaaaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAAT 180

Qy     181 aaaggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttg 240
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     181 AAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTG 240

Qy     241 gtcggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC 300

Qy     301 gtggcggcattggtgggcg-atcaatatattgtgagcgtggcacataacggcggctataa 359
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     301 GTGGCGGCATTGGTGGGCGTATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAA 360

Qy     360 caacgttgattttggtgcggaaggaagnaat---cccgatcagcaccgtttttcttacca 416
           |||||||||||||||||||||||||||||:||||   || ||||| || ||::|| |||| |
Db     361 CAACGTTGATTTTGGTGCGGAAGGAAKNAATATCCCNGATCAACAWCGWWTTACTTATAA 420

Qy     417 aattgtgaaaagaaataattataagcctgacaattca---caccccttacaacggcgatta 473
           |||||||||| | |||||||||||  | |  | | |   || |||||   |||||||||
Db     421 AATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTA 480

Qy     474 ncatatgccgcgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtga 533
           ||||||||||||||||||||:|||||||||||||||||||||| |||||||| ||| |
Db     481 TCATATGCCGCGTTTGCATAAATWTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTA 540

Qy     534 catgaggggggaatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctc 593
           |||   |||   |  |||  |||| || | || || || || ||||| || || ||  |
Db     541 TATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGC 600

Qy     594 aggacaccactattggcgttatgatgatgacaaacacggcgatttatcctactccggcgc 653
           |||    || |||||||| | ||||||  ||  | | |
Db     601 AGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCA---------------------- 638

Qy     654 atggttaattggcggcaatacacatatgcagggttggggaaataatggcgtanttagttt 713
```


Db 639 ATAACCGCGAAAGTTCATATCATATTGCAAGT 670
Qy 714 gagcggcgatgtgcgccatgccaacgactatggccctatgccgattgcaggtgcggcagg 773
Db 671 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 708
Qy 774 cgacagcggttcgccaatgtttatttatgacaaaacaaacaataaatggctgctcaacgg 833
Db 709 NNNNNNNGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGG 768
Qy 834 agttttacaaaccggctacccttattccggcagggaaaacggtttccagctgatacgcaa 893
Db 769 GGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAA 828
Qy 894 agattggttctacgatgacatttacagaggcgatacacataccgtctntttgaaccgcg 953
Db 829 AGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 888
Qy 954 cagtaacggacatttttcctttacatccaacaacaacggtacgggtacggtaacagaaac 1013
Db 889 TCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAA 948
Qy 1014 caacgaaaaggtntccaatccaaa---gcttaaagtacagacagtccgactgtttgacga 1070
Db 949 ACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1008
Qy 1071 atctttgaatgaaactgataaagaaccagttt---acgcggcaggggtgttaatcagta 1127
Db 1009 TTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTA 1068
Qy 1128 ccgtccaaggttaaacaacggtgaaaacctttcttttatcgattacggcaacggcaaact 1187
Db 1069 TCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATT 1128
Qy 1188 catcttatcaaacaacatcaaccaaggcgcgggcggtttgtattttgaaggtgatttttac 1247
Db 1129 GATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTTAC 1188
Qy 1248 ggtctcgcctgaaaacaacgaaacgtggcaaggcgcgggcgttcatatcagtgaagacag 1307
Db 1189 GGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAG 1248
Qy 1308 taccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtccaaaatcggcaaagg 1367
Db 1249 TACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGG 1308
Qy 1368 cacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagcgtgggcgacggtac 1427
Db 1309 CACGCTG---------NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1359
Qy 1428 agtcattttggatcagcaggcagacgataaaggcaaaaaacaagcctttagtgaaatcgg 1487
Db 1360 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1419
Qy 1488 cttgntcagcggcaggggtacggtgcaactgaatgccgataatcagttcaaccccgacaa 1547
Db 1420 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1479
Qy 1548 actctatttcggctttcgcggcggacgtttggatttaaacgggcattcgctttcgttcca 1607
Db 1480 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1539
Qy 1608 ccgtattcaaaataccgatgaaggggcgatgattgncnatcataatgccacaacaacatc 1667
Db 1540 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1599
Qy 1668 caccgttaccattacagggaatgaaagtattacacaaccgagtggtaagaatatcaatag 1727
Db 1600 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1659
Qy 1728 acttaattacagcaaagaaattgcctacaacggttggtttggcgagaaagatacgaccaa 1787
Db 1660 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1719
Qy 1788 aacgaacgggcggctcaaccttgtttaccagcccgccgcagaagaccgcacccngctgct 1847
Db 1720 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1779
Qy 1848 ttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggcaaactgtttttcag 1907
Db 1780 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1839
Qy 1908 cggcagaccgacaccgcacgcctacaatcatttaggaagcgggtggtcaaaaatggaagg 1967
Db 1840 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1899
Qy 1968 tatcccacaaggagaaatcgtgtgggacaacgactggatcnaccgcacgtttaaagcgga 2027
Db 1900 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1959
Qy 2028 aaatttccatattcagggcgggcaggcggtgatttcccgcaatgttgccaaagtggaagg 2087
Db 1960 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2019
Qy 2088 cgattgncatttgagcaatcacgcccaagcagtttttggtgtcgcaccgcatcaaagcca 2147
Db 2020 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2079
Qy 2148 tacaatctgtacacgttcggactggacnggtctgacaaattgtgtcgaanaaancattac 2207
Db 2080 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2139
Qy 2208 cgacgataaagtgattgcttcattgactaagacngacntnagcggcantgtnagnctnnc 2267
Db 2140 NNNNGATAAAGTGACTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGC 2199
Qy 2268 cnatnacgntnnttnaaanctcncngggcntgcnncactnaanggcaatcttagtgcaaa 2327
Db 2200 CGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAA 2259
Qy 2328 tggcgatacacgttatacagtcagccacaacgccacccaaaacggcaaccttagcctcgt 2387
Db 2260 TGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCKTAGCCTCGT 2319
Qy 2388 gggcaatgcccaagcaacatttaatcaagccacattaaacggcaacncatcggnttcggg 2447
Db 2320 GGNSAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGG 2379
Qy 2448 caatgcttcatttaatctaagcaacaacgccgcacaaaacggcagtctgacgctttccga 2507
Db 2380 CAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGG 2439
Qy 2508 caacgctaaggcaaacgtaagccattccgcactcaacggcaatgtctccctagccgataa 2567
Db 2440 CAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAA 2499
Qy 2568 ggcagtattccattttgaaaacagccgctttaccggacaactcagcggcagcaagganac 2627
Db 2500 GGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATAC 2559
Qy 2628 agcattacacttaaaagacagcgaatggacgctgccgtcaggcacggaattaggcaattt 2687
Db 2560 GGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGARCGGAATTAGGCAATTT 2619
Qy 2688 aaaccttgacaacgccaccattacactcaattccgcctatcgccacgatgctgcaggcgc 2747
Db 2620 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGC 2679
Qy 2748 gcaaaccggcagngtgtcagacacgccgcgc---------cgccgttcgcgccgttccct 2798
Db 2680 GCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCT 2739
Qy 2799 attatccgttacaccgccaacttcggtagaatcccgtttcaacacgctgacggtaaacgg 2858
Db 2740 ATTATMCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGG 2799

```
Qy   2859 caaattgaacngtcaaggaacattccgctttatgtcggaactcttcggctaccgaagcga 2918
          |||||||||| |||| ||||||||||||||||||||||||||||||||||||||||| |||||
Db   2800 CAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGA 2859

Qy   2919 caaattgaagctggcggaaagttccgaaggnacttacaccttggcggtcaacaataccgg 2978
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   2860 CAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGG 2919

Qy   2979 caacgaacccgtaagcctcgatcaattgacggtagtggaagggaaagacaacaaaccgct 3038
          ||||||||| | |||||||| |||||||||||||||||||| ||||||||||||||||||
Db   2920 CAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCT 2979

Qy   3039 gtccgaaaaccttaatttcaccctgcaaaacgaacacgtcgatgccggcgcgtgg----- 3093
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   2980 GTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCAGGCGCGTGGNNNNN 3039

Qy   3094 ----------------------------------------------------------cg 3095

Db   3040 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3099

Qy   3096 ttaccaactcatccgcaaagacggcgagttccgcctgcataatccggtcaaagaacaaga 3155

Db   3100 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3159

Qy   3156 gctttccgacaaactcggcaaggcagaagccaaaaaacaggcggaaaaagacaacgcgca 3215

Db   3160 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3219

Qy   3216 aagccttgacgcgctgattgcggccgggcgcgatgccgccgaaaagacagaaagcgttgc 3275

Db   3220 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3279

Qy   3276 cgaaccggcccggcnggcaggcggggaaaatgtcggcattatgcaggcggaggaagagaa 3335

Db   3280 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3339

Qy   3336 aaaacgggtgcaggcggataaagacagcgcnttggcgaaacagcgcgaagcggaaacccg 3395

Db   3340 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3399

Qy   3396 gccggntaccaccgccttcccccgcgcccgcngcgcccgccgggatttgccgcaaccgca 3455

Db   3400 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3459

Qy   3456 gccccaaccgcaacctcaaccccaaccgcagcgcgacctgatnagccgttatgccaatag 3515

Db   3460 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3519

Qy   3516 cggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacaggacgaattgga 3575
                                                                || ||
Db   3520 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTAGA 3579

Qy   3576 ccgcgtgtttgccgaagaccgccgcaacgcngtttggacaagcngcatccggnacaccaa 3635
          |||||| |||||||||||||||||||||| |||||||||||| |||||||| |||||||
Db   3580 CCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAA 3639

Qy   3636 acactaccgttcgcaagatttccgcgcctaccgccaacaaaccgacctgcgccaaatcgg 3695
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3640 ACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGG 3699

Qy   3696 tatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcacaaccggaccga 3755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3700 TATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGA 3759

Qy   3756 aaacancttcgacgacggcatcggcaactcggcacggcttgcccacggcgccgttttcgg 3815
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3760 AAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGG 3819

Qy   3816 gcaatacggcatcggcaggttcgacatcggcatcag-cacgggcgcgggttttagcagcg 3874
          |||||||||||||| ||||||| |||||||||||| | ||||||||||||||||||||
Db   3820 GCAATACGGCATCGACAGGTTCTACATCGGCATCAGNCGCGGGCGCGGGTTTTAGCAGCG 3879
```

```
Qy   3875 gcantctntcagacggcatcggaggcaaaatccgccgccgcgtgctgcattacggcattc 3934
          |||  || |||||||||||||||::|||:|||||||||||||||||||||||||||||||
Db   3880 GCAGCCTTTCAGACGGCATCGGAGSMAAAWTCCGCCGCCGCGTGCTGCATTACGGCATTC 3939

Qy   3935 aggcacgataccgcgccggtttcggcggattcggcatcgaaccgtacatcggcgcaacgc 3994
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   3940 AGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGC 3999

Qy   3995 gctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacccccggtc 4054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   4000 GCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCC 4059

Qy   4055 ttgcgttcaaccgntaccgngcgggcattaaggcagattattcattcaaaccggcgcaac 4114
          |||| |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db   4060 TTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAAC 4119

Qy   4115 acatntccatcacnccttatttnagcctgtcctataccgatgccgcttcgggcaaagtcc 4174
          |||| |||||||| |||||||| ||||||||||||||||||||||||||||||||||||
Db   4120 ACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCC 4179

Qy   4175 gaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgcagtgcggaat 4234
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4180 GAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAAT 4239

Qy   4235 ggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgccgccgccaaag 4294
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   4240 GGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAG 4299

Qy   4295 gnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctggtaa 4350
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4300 GCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGGTAA 4355
```

```
RESULT  12
HIU11024
LOCUS       HIU11024                4317 bp    DNA     linear   BCT 12-JUL-1996
DEFINITION  Haemophilus influenzae N187 adhesion and penetration protein (hap)
            gene, complete cds.
ACCESSION   U11024
VERSION     U11024.1  GI:1408568
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 4317)
  AUTHORS   St. Geme,J.W. III, de la Morena,M.L. and Falkow,S.
  TITLE     A Haemophilus influenzae IgA protease-like protein promotes
            intimate interaction with human epithelial cells
  JOURNAL   Mol. Microbiol. 14 (2), 217-233 (1994)
  MEDLINE   95131744
REFERENCE   2  (bases 1 to 4317)
  AUTHORS   St. Geme,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1994) Joseph W. St. Geme, Molecular Microbiology,
            Washington University School of Medicine, 660 South Euclid Ave.,
            St. Louis, MO 63110, USA
COMMENT     On Jul 12, 1996 this sequence version replaced gi:562013.
FEATURES             Location/Qualifiers
     source          1. .4317
                     /organism="Haemophilus influenzae"
                     /strain="N187"
                     /db_xref="taxon:727"
     gene            61. .4242
                     /gene="hap"
     CDS             61. .4242
                     /gene="hap"
                     /standard_name="Haemophilus adhesion and penetration"
                     /note="HAP"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesion and penetration protein"
                     /protein_id="AAB03707.1"
```

```
                     /db_xref="GI:1408569"
                     /translation="MKKTVFRLNFLTACISLGIVSQAWAGHTYFGIDYQYYRDFAENK
                     GKFTVGAQNIKVYNKQGQLVGTSMTKAPMIDFSVVSRNGVAALVENQYIVSVAHNVGY
                     TDVDFGAEGNNPDQHRFTYKIVKRNNYKKDNLHPYEDDYHNPRLHKFVTEAAPIDMTS
                     NMNGSTYSDRTKYPERVRIGSGRQFWRNDQDKGDQVAGAYHYLTAGNTHNQRGAGNGY
                     SYLGGDVRKAGEYGPLPIAGSKGDSGSPMFIYDAEKQKWLINGILREGNPFEGKENGF
                     QLVRKSYFDEIFERDLHTSLYTRAGNGVYTISGNDNGQGSITQKSGIPSEIKITLANM
                     SLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDQKQGSLIFASDINQGAGGLYFEG
                     NFTVSPNSNQTWQGAGIHVSENSTVTWKVNGVEHDRLSKIGKGTLHVQAKGENKGSIS
                     VGDGKVILEQQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKFYFGFRGGRLDLN
                     GHSLTFKRIQNTDEGAMIVNHNTTQAANVTITGNESIVLPNGNNINKLDYRKEIAYNG
                     WFGETDKNKHNGRLNLIYKPTTEDRTLLLSGGTNLKGDITQTKGKLFFSGRPTPHAYN
                     HLNKRWSEMEGIPQGEIVWDHDWINRTFKAENFQIKGGSAVVSRNVSSIEGNWTVSNN
                     ANATFGVVPNQQNTICTRSDWTGLTTCQKVDLTDTKVINSIPKTQINGSINLTDNATA
                     NVKGLAKLNGNVTLTNHSQFTLSNNATQIGNIRLSDNSTATVDNANLNGNVHLTDSAQ
                     FSLKNSHFSHQIQGDKGTTVTLENATWTMPSDTTLQNLTLNNSTITLNSAYSASSNNT
                     PRRRSLETETTPTSAEHRFNTLTVNGKLSGQGTFQFTSSLFGYKSDKLKLSNDAEGDY
                     ILSVRNTGKEPETLEQLTLVESKDNQPLSDKLKFTLENDHVDAGALRYKLVKNDGEFR
                     LHNPIKEQELHNDLVRAEQAERTLEAKQVEPTAKTQTGEPKVRSRRAARAAFPDTLPD
                     QSLLNALEAKQAELTAETQKSKAKTKKVRSKRAVFSDPLLDQSLFALEAALEVIDAPQ
                     QSEKDRLAQEEAEKQRKQKDLISRYSNSALSELSATVNSMLSVQDELDRLFVDQAQSA
                     VWTNIAQDKRRYDSDAFRAYQQKTNLRQIGVQKALANGRIGAVFSHSRSDNTFDEQVK
                     NHATLTMMSGFAQYQWGDLQFGVNVGTGISASKMAEEQSRKIHRKAINYGVNASYQFR
                     LGQLGIQPYFGVNRYFIERENYQSEEVRVKTPSLAFNRYNAGIRVDYTFTPTDNISVK
                     PYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKSQGSQ
                     LGKQQNVGVKLGYRW"
BASE COUNT     1496 a    776 c    890 g   1155 t
ORIGIN


  Query Match           31.3%;  Score 1359.4;  DB 1;  Length 4317;
  Best Local Similarity 59.0%;  Pred. No. 3.4e-300;
  Matches 2550;  Conservative    0;  Mismatches 1566;  Indels  207;  Gaps    8;

Qy      73 cctgcttacttagccatatgcctgtcgttcggcattcttccccaagcttgggcgggacac 132
           | |  ||  ||| ||   ||| | || || || ||  |  | |||||| |||||| || |||
Db      82 CTTAATTTTTTAACCGCTTGCATTTCATTAGGGATAGTATCGCAAGCGTGGGCTGGTCAC 141

Qy     133 acttatttcggcatcaactaccaatactatcgcgactttgccgaaaataaaggcaagttt 192
           |||||||| || ||  | |||||||| ||||| || |||||||| |||||||| |||||
Db     142 ACTTATTTTGGGATTGATTACCAATATTATCGTGATTTTGCCGAGAATAAAGGGAAGTTC 201

Qy     193 gcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttggtcggcaaatca 252
           |||| |||||  || ||||| |||| || |||||| ||||| | || || |||| ||||
Db     202 ACAGTTGGGGCTCAAAATATTAAGGTTTATAACAAACAAGGGCAATTAGTTGGCACATCA 261

Qy     253 atgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattg 312
           |||||||||||||||||||||||||||||||||| ||||| |||||||||||||| || |||
Db     262 ATGACAAAAGCCCCGATGATTGATTTTTCTGTAGTGTCACGTAACGGCGTGGCAGCCTTG 321

Qy     313 gtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgatttt 372
           || |   |||||||||||||||||||||||||||||  || ||||   | |||||||||
Db     322 GTTGAAAATCAATATATTGTGAGCGTGGCACATAACGTAGGATATACAGATGTTGATTTT 381

Qy     373 ggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaattgtgaaaagaaat 432
           ||||| || ||||  || |||||||| || |||||| |||| | ||||| ||| ||||| 
Db     382 GGTGCAGAGGGAAACAACCCCGATCAACATCGTTTTACTTATAAGATTGTAAAACGAAAT 441

Qy     433 aattataagcctgacaattcacacccttacaacggcgattancatatgccgcgtttgcat 492
           || || ||    || |||| |||| |||||  | | |||||| |||| || || || |||
Db     442 AACTACAAAAAAGATAATTTACATCCTTATGAGGACGATTACCATAATCCACGATTACAT 501

Qy     493 aaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgaggggaatacctat 552
           ||||| || ||||| || |  ||  | || |||||     | ||||  || | ||| |||
Db     502 AAATTCGTTACAGAAGCGGCTCCAATTGATATGACTTCGAATATGAATGGCAGTACTTAT 561

Qy     553 tccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacaccactattggcgt 612
           || |||| |  |||||||| || ||||| || |||||||| |||| || | ||||||
Db     562 TCAGATAGAACAAAATATCCAGAACGTGTTCGTATCGGCTCTGGACGGCAGTTTTGGCGA 621

Qy     613 tatgatgatgacaaacacggcgatttatcctactccggcgcatggttaattggcggcaat 672
           |||||  | ||||||  || | |  |  ||     |       |  | |  |  ||||||
Db     622 AATGATCAAGACAAAGGCGACCAAGTTGCCGGTGCATATCATTATCTGACAGCTGGCAAT 681

Qy     673 acacatatgcagggttggggaaataatggcgtanttagtttgagcggcgatgtgcgccat 732
           ||||| |  ||| || | | |  ||||||         |||| | |||||||| ||  |
Db     682 ACACACAATCAGCGTGGAGCAGGTAATGGATATTCGTATTTGGGAGGCGATGTTCGTAAA 741

Qy     733 gccaacgactatggccctatgccgattgcaggtgcggcaggcgacagcggttcgccaatg 792
           ||    || ||||| ||  | ||||||||||| |    || ||||| ||||| || |||
Db     742 GCGGGAGAATATGGTCCATTACCGATTGCAGGCTCAAAGGGGGACAGTGGTTCTCCGATG 801

Qy     793 tttatttatgacaaaacaaacaataaatggctgctcaacggagttttacaaaccggctac 852
           ||||||||||      ||| | |||||| |  | || ||  | ||||     ||| ||
Db     802 TTTATTTATGATGCTGAAAAACAAAAATGGTTAATTAATGGGATATTACGGGAAGGCAAC 861

Qy     853 ccttattccggcagggaaaacggtttccagctgatacgcaaagattggttctacgatgac 912
           |||| |   ||||  ||||| || || ||  || | ||||||  ||  ||    |||||
Db     862 CCTTTTGAAGGCAAAGAAAATGGGTTTCAATTGGTTCGCAAATCTTATTT---TGATGAA 918

Qy     913 atttacagaggcgatacacataccgtctntttgaaccgcgcagtaacggacatttttcc 972
           |||| |  | | ||| ||||||      ||      |     |||| ||| |  |  |
Db     919 ATTTTCGAAAGAGATTTACATACATCACTTTACACCCGAGCTGGTAATGGAGTGTACACA 978

Qy     973 tttacatccaacaacaacggtacgggtacggtaacagaaaccaacgaaaaggtntccaat 1032
           |||     ||  | || |||  |||  |  ||||  | |     | ||     ||  |
Db     979 ATTAGTGGAAATGATAATGGTCAGGGGTCTATAACTCAGAAATCAGGAATACCATCAGAA 1038

Qy    1033 ccaaagcttaaagtacagacagtccgactgtttgacgaatctttgaatgaaactgataaa 1092
              ||  |||   ||  |  |  |  |        | ||||||| || |  ||||||
Db    1039 ATTAAAATTACGTTAGCAAATATGAGTTT---------ACCTTTGAAAGAGAAGGATAAA 1089

Qy    1093 gaaccagtttacgcggcaggggggtgttaatcagtaccgtccaaggttaaacaacggtgaa 1152
           |  |  |       |  ||  ||||  ||  |||| | |||||||| || |||
Db    1090 GTTCATAATCCTAGATATGACGGACCTAATATTTATTCTCCACGTTTAAACAATGGAGAA 1149

Qy    1153 aacctttcttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaa 1212
           |  || | |||||| ||| |     | ||   | | |||||  ||    |||| ||||||
Db    1150 ACGCTATATTTTATGGATCAAAAACAAGGATCATTAATCTTCGCATCTGACATTAACCAA 1209

Qy    1213 ggcgcgggcggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacgaaacg 1272
           || ||||| ||| | |||||||| ||| ||||||| || || ||  |    ||| ||||
Db    1210 GGGGCGGGTGGTCTTTATTTTGAGGGTAATTTTACAGTATCTCCAAATTCTAACCAAACT 1269

Qy    1273 tggcaaggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggc 1332
           |||||||| || ||| | ||| | |||||| | || ||||||||||||||||||||| |||
Db    1270 TGGCAAGGAGCTGGCATACATGTAAGTGAAAATAGCACCGTTACTTGGAAAGTAAATGGC 1329

Qy    1333 gtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggg 1392
           |||| | | || || || || ||||| || ||||| ||  ||||||||||||||||||||
Db    1330 GTGGAACATGATCGACTTTCTAAAATTGGTAAAGGAACATTGCACGTTCAAGCCAAAGGG 1389

Qy    1393 gaaaaccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagac 1452
           |||||  |||| |||||||||||| ||||| |||| ||||||||||||| ||||||||||||
Db    1390 GAAAATAAAGGTTCGATCAGCGTAGGCGATGGTAAAGTCATTTTGGAGCAGCAGGCAGAC 1449

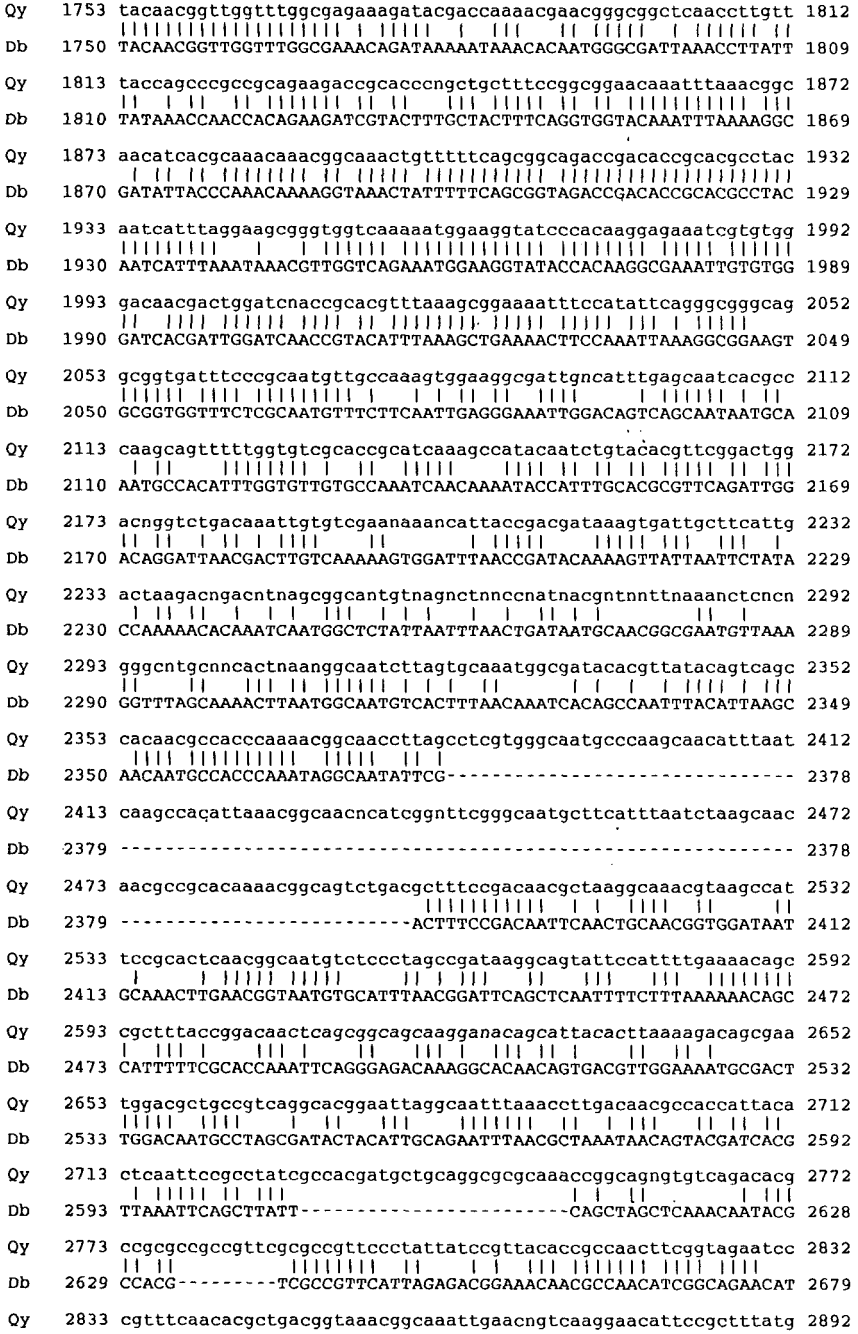
Qy    1453 gataaaggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtg 1512
           ||| ||||||| |||||||||||||||||||||| |||||| | |||||||| || || ||
Db    1450 GATCAAGGCAACAAACAAGCCTTTAGTGAAATTGGCTTGGTTAGCGGCAGAGGGACTGTT 1509

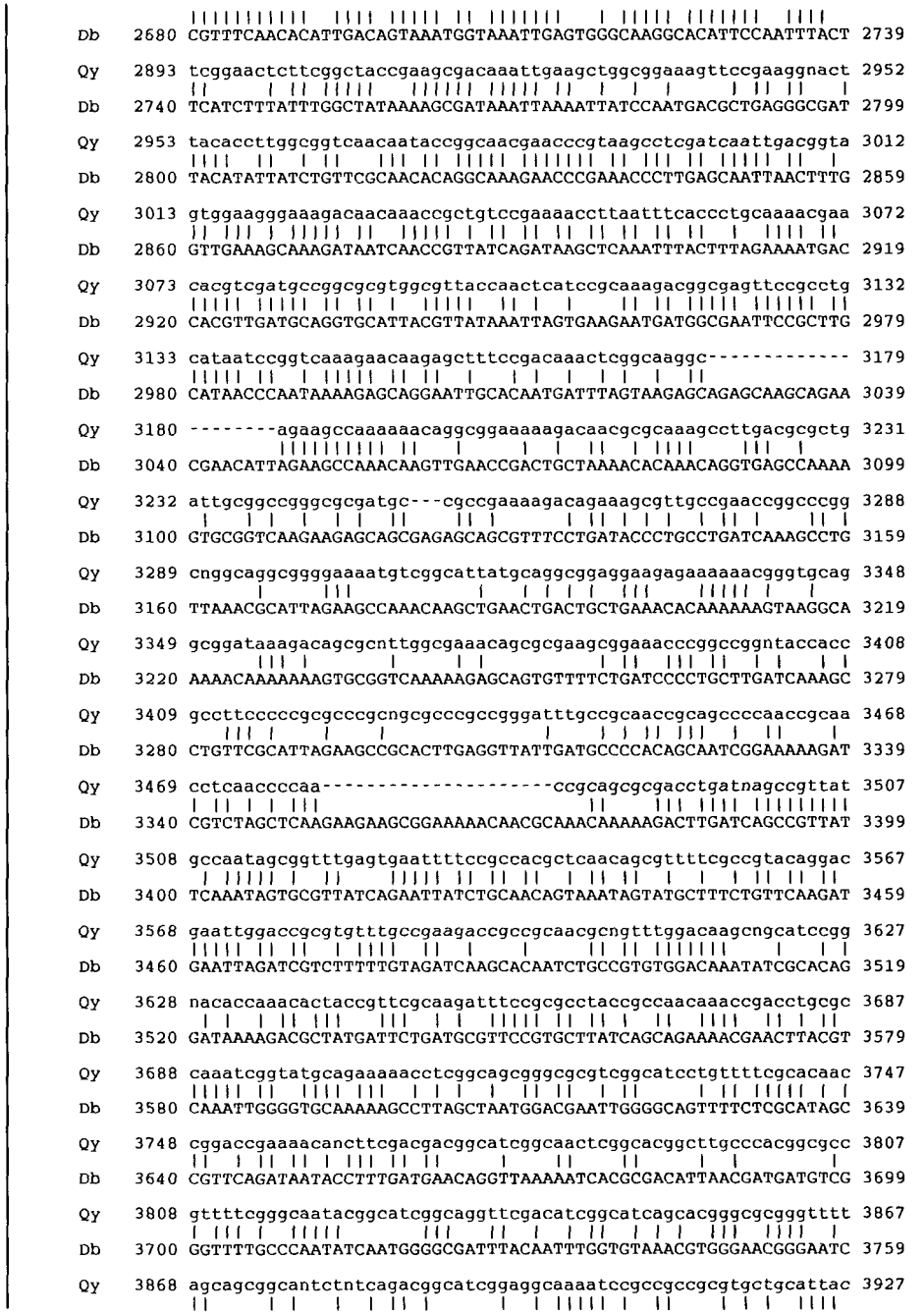
Qy    1513 caactgaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcgga 1572
           ||| | || |  ||||| || ||  |  |||| ||| | |||||||||||||| || ||
Db    1510 CAATTAAACGATGATAAACAATTTGATACCGATAAATTTTATTTCGGCTTTCGTGGTGGT 1569

Qy    1573 cgtttggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggg 1632
           || || ||| | |||||||||||  |  | ||  | ||||| |||||||| || || |||
Db    1570 CGCTTAGATCTTAACGGGCATTCATTAACCTTTAAACGTATCCAAAATACGGACGAGGGG 1629

Qy    1633 gcgatgattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaa 1692
           || |||||||   | |||||| | ||   | |  | |  || || ||||| |||| |||
Db    1630 GCAATGATTGTGAACCATAATACAACTCAAGCCGCTAATGTCACTATTACTGGGAACGAA 1689

Qy    1693 agtattacacaaccgagtggtaagaatatcaatagacttaattacagcaaagaaattgcc 1752
           || |||   | ||| | ||| || ||||| |||| |||| ||||||| ||||||||||||
Db    1690 AGCATTGTTCTACCTAATGGAAATAATATTAATAAACTTGATTACAGAAAAGAAATTGCC 1749
```





```
Db   3760 AGTGCGAGTAAAATGGCTGAAGAACAAAGCCGAAAAATTCATCGAAAAGCGATAAATTAT 3819

Qy   3928 ggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgtacatcggc 3987
          ||| |  | ||| | || |    | |||| ||   ||| |||||  | || ||  | ||
Db   3820 GGCGTGAATGCAAGTTATCAGTTCCGTTTAGGGCAATTGGGCATTCAGCCTTATTTTGGA 3879

Qy   3988 gcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgccacc 4047
          |  |  ||||||||  |  ||   |  |||| |  |  ||  | || |   |    ||
Db   3880 GTTAATCGCTATTTTATTGAACGTGAAAATTATCAATCTGAGGAAGTGAGAGTGAAAACG 3939

Qy   4048 cccggtcttgcgttcaaccgntaccgngcgggcattaaggcagattattcattcaaaccg 4107
          ||  | ||||| || || || ||    || ||||||   |  |||||| |||| |  |||
Db   3940 CCTAGCCTTGCATTTAATCGCTATAATGCTGGCATTCGAGTTGATTATACATTTACTCCG 3999

Qy   4108 gcgcaacacatntccatcacnccttatttnagcctgtcctataccgatgccgcttcgggc 4167
           |  |  | ||   | | |  |||||||||   | |    |||   ||||   |    |
Db   4000 ACAGATAATATCAGCGTTAAGCCTTATTTCTTCGTCAATTATGTTGATGTTTCAAACGCT 4059

Qy   4168 aaagtccgaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgcagt 4227
          || || | |||    || |||  | | || |||   ||   || ||        |
Db   4060 AACGTACAAACCACGGTAAATCTCACGGTGTTGCAACAACCATTTGGACGTTATTGGCAA 4119

Qy   4228 gcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgccgcc 4287
           |||  |||  |||| || |||||   |  ||||   | |||          |
Db   4120 AAAGAAGTGGGATTAAAGGCAGAAATTTTACATTTCCAAATTTCCGCTTTTATCTCAAAA 4179

Qy   4288 gccaaaggnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgctgg 4347
           |  ||||  | ||||| |     || || |  | |||| | ||||| ||||| || |||
Db   4180 TCTCAAGGTTCACAACTCGGCAAACAGCAAAATGTGGGCGTGAAATTGGGCTATCGTTGG 4239

Qy   4348 taa 4350
          |||
Db   4240 TAA 4242


RESULT  13
AR157320
LOCUS       AR157320                4319 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6245337.
ACCESSION   AR157320
VERSION     AR157320.1  GI:16218252
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4319)
  AUTHORS   St. Geme,J.W. III and Falkow,S.
  TITLE     Haemophilus adherence and penetration proteins
  JOURNAL   Patent: US 6245337-A 1 12-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .4319
                     /organism="unknown"
BASE COUNT     1497 a    776 c    891 g   1155 t
ORIGIN


  Query Match             31.0%;  Score 1348;  DB 6;  Length 4319;
  Best Local Similarity   59.0%;  Pred. No. 1.4e-297;
  Matches 2551;  Conservative    0;  Mismatches 1565;  Indels  210;  Gaps    9;

Qy     73 cctgcttacttagccatatgcctgtcgttcggcattcttccccaagcttgggcgggacac 132
          | |  ||  ||| ||   ||| | || || || ||  |  | ||||| ||||| || |||
Db     81 CTTAATTTTTTAACCGCTTGCATTTCATTAGGGATAGTATCGCAAGCGTGGGCTGGTCAC 140

Qy    133 acttatttcggcatcaactaccaatactatcgcgactttgccgaaaataaaggcaagttt 192
          ||||||||  || ||  | |||||||| ||||| || |||||||| |||||||| |||||
Db    141 ACTTATTTTGGGATTGATTACCAATATTATCGTGATTTTGCCGAGAATAAAGGGAAGTTC 200

Qy    193 gcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttggtcggcaaatca 252
          ||||  |||||  || ||||| |||| || |||||| ||||| | || || |||| ||||
Db    201 ACAGTTGGGGCTCAAAATATTAAGGTTTATAACAAACAAGGGCAATTAGTTGGCACATCA 260

Qy    253 atgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattg 312
          |||||||||||||||||||||||||||||||||||||| ||||| |||||||||||||| || |||
Db    261 ATGACAAAAGCCCCGATGATTGATTTTTCTGTAGTGTCACGTAACGGCGTGGCAGCCTTG 320

Qy    313 gtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgatttt 372
          || |  |||||||||||||||||||||||||||||  || ||||  | |||||||||
Db    321 GTTGAAAATCAATATATTGTGAGCGTGGCACATAACGTAGGATATACAGATGTTGATTTT 380

Qy    373 ggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaattgtgaaaagaaat 432
          ||||| || |||| || |||||||| || |||||| |||| | ||||| ||| |||||
Db    381 GGTGCAGAGGGAAACAACCCCGATCAACATCGTTTTACTTATAAGATTGTAAAACGAAAT 440

Qy    433 aattataagcctgacaattcacacccttacaacggcgattancatatgccgcgtttgcat 492
          || || ||   || |||| ||| ||||| | | |||||| |||| || || || |||
Db    441 AACTACAAAAAAGATAATTTACATCCTTATGAGGACGATTACCATAATCCACGATTACAT 500

Qy    493 aaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgaggggaatacctat 552
          ||||| || ||||| || |  ||  | || ||||||     | ||||  || | ||| |||
Db    501 AAATTCGTTACAGAAGCGGCTCCAATTGATATGACTTCGAATATGAATGGCAGTACTTAT 560

Qy    553 tccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacaccactattggcgt 612
          || |||| |  |||||||| || |||| || |||||||| |||| || | ||||||
Db    561 TCAGATAGAACAAAATATCCAGAACGTGTTCGTATCGGCTCTGGACGGCAGTTTTGGCGA 620

Qy    613 tatgatgatgacaaacacggcgatttatcctactccggcgcatggttaattggcggcaat 672
          |||||  | ||||||  || | |  |  ||   |      |   | |  |  ||||||
Db    621 AATGATCAAGACAAAGGCGACCAAGTTGCCGGTGCATATCATTATCTGACAGCTGGCAAT 680

Qy    673 acacatatgcagggttggggaaataatggcgtanttagtttgagcggcgatgtgcgccat 732
          |||||  |  ||| || | | |  ||||||          |||| | |||||||| ||  |
Db    681 ACACACAATCAGCGTGGAGCAGGTAATGGATATTCGTATTTGGGAGGCGATGTTCGTAAA 740

Qy    733 gccaacgactatggccctatgccgattgcaggtgcggcaggcgacagcggttcgccaatg 792
          ||    || ||||| || | ||||||||||  |    || ||||| ||||| || |||
Db    741 GCGGGAGAATATGGTCCATTACCGATTGCAGGCTCAAAGGGGGACAGTGGTTCTCCGATG 800

Qy    793 tttatttatgacaaaacaaacaataaatggctgctcaacggagttttacaaaccggctac 852
          ||||||||||       |||  | ||||||  |  | || ||  | ||||     ||| ||
Db    801 TTTATTTATGATGCTGAAAAACAAAAATGGTTAATTAATGGGATATTACGGGAAGGCAAC 860

Qy    853 ccttattccggcagggaaaacggtttccagctgatacgcaaagattggttctacgatgac 912
          |||| |   ||||  ||||| || || ||  || | ||||||  ||  ||    |||||
Db    861 CCTTTTGAAGGCAAAGAAAATGGGTTTCAATTGGTTCGCAAATCTTATTT---TGATGAA 917

Qy    913 atttacagaggcgatacacataccgtctnttttgaaccgcgcagtaacggacatttttcc 972
          |||| |  | | ||| ||||||      ||      |     |||| |||   |   |
Db    918 ATTTTCGAAAGAGATTTACATACATCACTTTACACCCGAGCTGGTAATGGAGTGTACACA 977

Qy    973 tttacatccaacaacaacggtacgggtacggtaacagaaaccaacgaaaaggtntccaat 1032
          |||     ||  | || |||  |||  |  ||||  | |     | ||    ||  |
Db    978 ATTAGTGGAAATGATAATGGTCAGGGGTCTATAACTCAGAAATCAGGAATACCATCAGAA 1037

Qy   1033 ccaaagcttaaagtacagacagtccgactgtttgacgaatctttgaatgaaactgataaa 1092
            ||  |||   ||   |  |  | |          | ||||||| || |  ||||||
Db   1038 ATTAAAATTACGTTAGCAAATATGAGTTT---------ACCTTTGAAAGAGAAGGATAAA 1088

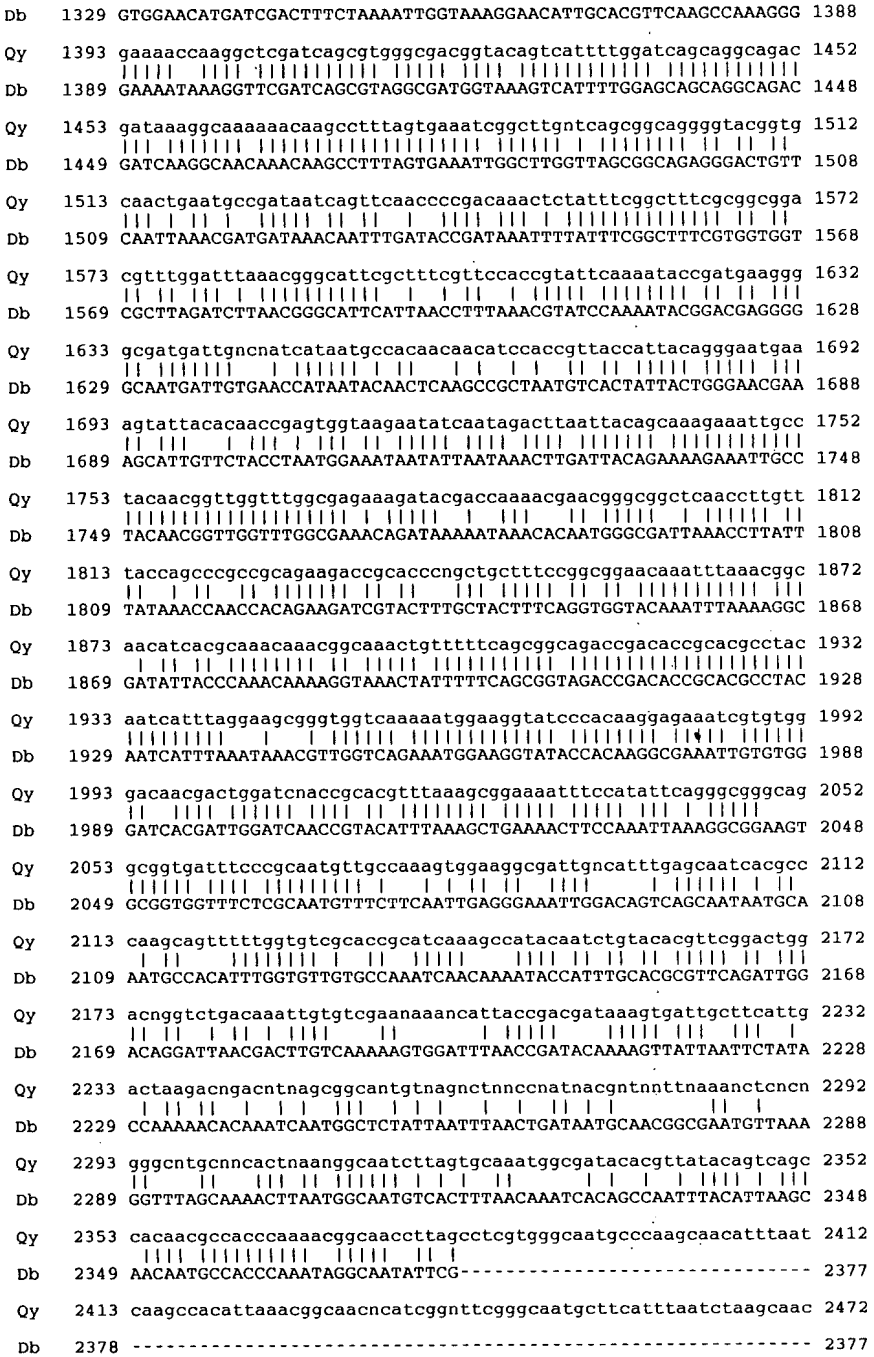
Qy   1093 gaaccagtttacgcggcaggggtgttaatcagtaccgtccaaggttaaacaacggtgaa 1152
          |  |   |         |  ||  ||||  ||  |||| |  ||||||| || |||
Db   1089 GTTCATAATCCTAGATATGACGGACCTAATATTTATTCTCCACGTTTAAACAATGGAGAA 1148

Qy   1153 aacctttcttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaa 1212
          |  || | ||||||  ||| |    | ||   | | |||||  ||    |||| ||||||
Db   1149 ACGCTATATTTTATGGATCAAAAACAAGGATCATTAATCTTCGCATCTGACATTAACCAA 1208

Qy   1213 ggcgcgggcggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacgaaacg 1272
          || ||||| ||| | |||||||| ||| ||||||| || || ||  |    ||| ||||
Db   1209 GGGGCGGGTGGTCTTTATTTTGAGGGTAATTTTACAGTATCTCCAAATTCTAACCAAACT 1268

Qy   1273 tggcaaggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggc 1332
          |||||||| || ||| | ||| | |||||| | || ||||||||||||||||||| |||
Db   1269 TGGCAAGGAGCTGGCATACATGTAAGTGAAAATAGCACCGTTACTTGGAAAGTAAATGGC 1328

Qy   1333 gtggcaaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggg 1392
          |||| | | || || || || ||||| || ||||| ||  ||||||||||||||||||||
```


Db 1329 GTGGAACATGATCGACTTTCTAAAATTGGTAAAGGAACATTGCACGTTCAAGCCAAAGGG 1388
Qy 1393 gaaaaccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagac 1452
Db 1389 GAAAATAAAGGTTCGATCAGCGTAGGCGATGGTAAAGTCATTTTGGAGCAGCAGGCAGAC 1448
Qy 1453 gataaaggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtg 1512
Db 1449 GATCAAGGCAACAAACAAGCCTTTAGTGAAATTGGCTTGGTTAGCGGCAGAGGGACTGTT 1508
Qy 1513 caactgaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcgga 1572
Db 1509 CAATTAAACGATGATAAACAATTTGATACCGATAAATTTTATTTCGGCTTTCGTGGTGGT 1568
Qy 1573 cgtttggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggg 1632
Db 1569 CGCTTAGATCTTAACGGGCATTCATTAACCTTTAAACGTATCCAAAATACGGACGAGGGG 1628
Qy 1633 gcgatgattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaa 1692
Db 1629 GCAATGATTGTGAACCATAATACAACTCAAGCCGCTAATGTCACTATTACTGGGAACGAA 1688
Qy 1693 agtattacacaaccgagtggtaagaatatcaatagacttaattacagcaaagaaattgcc 1752
Db 1689 AGCATTGTTCTACCTAATGGAAATAATATTAATAAACTTGATTACAGAAAAGAAATTGCC 1748
Qy 1753 tacaacggttggtttggcgagaaagatacgaccaaaacgaacgggcggctcaaccttgtt 1812
Db 1749 TACAACGGTTGGTTTGGCGAAACAGATAAAAATAAACACAATGGGCGATTAAACCTTATT 1808
Qy 1813 taccagcccgccgcagaagaccgcacccngctgctttccggcggaacaaatttaaacggc 1872
Db 1809 TATAAACCAACCACAGAAGATCGTACTTTGCTACTTTCAGGTGGTACAAATTTAAAAGGC 1868
Qy 1873 aacatcacgcaaacaaacggcaaactgtttttcagcggcagaccgacaccgcacgcctac 1932
Db 1869 GATATTACCCAAACAAAAGGTAAACTATTTTTCAGCGGTAGACCGACACCGCACGCCTAC 1928
Qy 1933 aatcatttaggaagcgggtggtcaaaaatggaaggtatcccacaaggagaaatcgtgtgg 1992
Db 1929 AATCATTTAAATAAACGTTGGTCAGAAATGGAAGGTATACCACAAGGCGAAATTGTGTGG 1988
Qy 1993 gacaacgactggatcnaccgcacgtttaaagcggaaaatttccatattcagggcgggcag 2052
Db 1989 GATCACGATTGGATCAACCGTACATTTAAAGCTGAAAACTTCCAAATTAAAGGCGGAAGT 2048
Qy 2053 gcggtgatttcccgcaatgttgccaaagtggaaggcgattgncatttgagcaatcacgcc 2112
Db 2049 GCGGTGGTTTCTCGCAATGTTTCTTCAATTGAGGGAAATTGGACAGTCAGCAATAATGCA 2108
Qy 2113 caagcagtttttggtgtcgcaccgcatcaaagccatacaatctgtacacgttcggactgg 2172
Db 2109 AATGCCACATTTGGTGTTGTGCCAAATCAACAAAATACCATTTGCACGCGTTCAGATTGG 2168
Qy 2173 acnggtctgacaaattgtgtcgaanaaancattaccgacgataaagtgattgcttcattg 2232
Db 2169 ACAGGATTAACGACTTGTCAAAAAGTGGATTTAACCGATACAAAAGTTATTAATTCTATA 2228
Qy 2233 actaagacngacntnagcggcantgtnagnctnnccnatnacgntnnttnaaanctcncn 2292
Db 2229 CCAAAAACACAAATCAATGGCTCTATTAATTTAACTGATAATGCAACGGCGAATGTTAAA 2288
Qy 2293 gggcntgcnncactnaanggcaatcttagtgcaaatggcgatacacgttatacagtcagc 2352
Db 2289 GGTTTAGCAAAACTTAATGGCAATGTCACTTTAACAAATCACAGCCAATTTACATTAAGC 2348
Qy 2353 cacaacgccacccaaaacggcaaccttagcctcgtgggcaatgcccaagcaacatttaat 2412
Db 2349 AACAATGCCACCCAAATAGGCAATATTCG---------- 2377
Qy 2413 caagccacattaaacggcaacncatcggnttcgggcaatgcttcatttaatctaagcaac 2472
Db 2378 ---------- 2377
Qy 2473 aacgccgcacaaaacggcagtctgacgctttccgacaacgctaaggcaaacgtaagccat 2532
Db 2378 ----------ACTTTCCGACAATTCAACTGCAACGGTGGATAAT 2411
Qy 2533 tccgcactcaacggcaatgtctccctagccgataaggcagtattccattttgaaaacagc 2592
Db 2412 GCAAACTTGAACGGTAATGTGCATTTAACGGATTCAGCTCAATTTTCTTTAAAAAACAGC 2471
Qy 2593 cgctttaccggacaactcagcggcagcaaggancacagcattacacttaaaagacagcgaa 2652
Db 2472 CATTTTTCGCACCAAATTCAGGGAGACAAAGGCACAACAGTGACGTTGGAAAATGCGACT 2531
Qy 2653 tggacgctgccgtcaggcacggaattaggcaatttaaaccttgacaacgccaccattaca 2712
Db 2532 TGGACAATGCCTAGCGATACTACATTGCAGAATTTAACGCTAAATAACAGTACGATCACG 2591
Qy 2713 ctcaattccgcctatcgccacgatgctgcaggcgcgcaaaccggcagngtgtcagacacg 2772
Db 2592 TTAAATTCAGCTTATT----------CAGCTAGCTCAAACAATACG 2627
Qy 2773 ccgcgccgccgttcgcgccgttccctattatccgttacaccgccaacttcggtagaatcc- 2832
Db 2628 CCACG---------TCGCCGTTCATTAGAGACGGAAACAACGCCAACATCGGCAGAACAT 2678
Qy 2833 cgtttcaacacgctgacggtaaacggcaaattgaacngtcaaggaacattccgctttatg 2892
Db 2679 CGTTTCAACACATTGACAGTAAATGGTAAATTGAGTGGGCAAGGCACATTCCAATTTACT 2738
Qy 2893 tcggaactcttcggctaccgaagcgacaaattgaagctggcggaaagttccgaaggnact 2952
Db 2739 TCATCTTTATTTGGCTATAAAAGCGATAAATTAAAATTATCCAATGACGCTGAGGGCGAT 2798
Qy 2953 tacaccttggcggtcaacaataccggcaacgaacccgtaagcctcgatcaattgacggta 3012
Db 2799 TACATATTATCTGTTCGCAACACAGGCAAAGAACCCGAAACCCTTGAGCAATTAACTTTG 2858
Qy 3013 gtggaagggaaagacaacaaaccgctgtccgaaaaccttaatttcaccctgcaaaacgaa 3072
Db 2859 GTTGAAAGCAAAGATAATCAACCGTTATCAGATAAGCTCAAATTTACTTTAGAAAATGAC 2918
Qy 3073 cacgtcgatgccggcgcgtggcgttaccaactcatccgcaaagacggcgagttccgcctg 3132
Db 2919 CACGTTGATGCAGGTGCATTACGTTATAAATTAGTGAAGAATGATGGCGAATTCCGCTTG 2978
Qy 3133 cataatccggtcaaagaacaagagctttccgacaaactcggcaaggc---------- 3179
Db 2979 CATAACCCAATAAAAGAGCAGGAATTGCACAATGATTTAGTAAGAGCAGAGCAAGCAGAA 3038
Qy 3180 --------agaagccaaaaaacaggcggaaaaagacaacgcgcaaagccttgacgcgctg 3231
Db 3039 CGAACATTAGAAGCCAAACAAGTTGAACCGACTGCTAAAACACAAACAGGTGAGCCAAAA 3098
Qy 3232 attgcggccgggcgcgatgc---cgccgaaaagacagaaagcgttgccgaaccggcccgg 3288
Db 3099 GTGCGGTCAAGAAGAGCAGCGAGAGCAGCGTTTCCTGATACCCTGCCTGATCAAAGCCTG 3158
Qy 3289 cnggcaggcggggaaaatgtcggcattatgcaggcggaggaagagaaaaaacgggtgcag 3348
Db 3159 TTAAACGCATTAGAAGCCAAACAAGCTGAACTGACTGCTGAAACACAAAAAAGTAAGGCA 3218
Qy 3349 gcggataaagacagcgcnttggcgaaacagcgcgaagcggaaacccggccggntaccacc 3408
Db 3219 AAAACAAAAAAAGTGCGGTCAAAAAGAGCAGTGTTTTCTGATCCCCTGCTTGATCAAAGC 3278
Qy 3409 gccttcccccgcgcccgcngcgcccgccgggatttgccgcaaccgcagccccaaccgcaa 3468
Db 3279 CTGTTCGCATTAGAAGCCGCACTTGAGGTTATTGATGCCCCACAGCAATCGGAAAAAGAT 3338
Qy 3469 cctcaaccccaa----------ccgcagcgcgacctgatnagccgttat 3507
Db 3339 CGTCTAGCTCAAGAAGAAGCGGAAAAACAACGCAAACAAAAAGACTTGATCAGCCGTTAT 3398

```
Qy   3508 gccaatagcggtttgagtgaattttccgccacgctcaacagcgttttcgccgtacaggac 3567
Db   3399 TCAAATAGTGCGTTATCAGAATTATCTGCAACAGTAAATAGTATGCTTTCTGTTCAAGAT 3458

Qy   3568 gaattggaccgcgtgtttgccgaagaccgccgcaacgcngtttggacaagcngcatccgg 3627
Db   3459 GAATTAGATCGTCTTTTTGTAGATCAAGCACAATCTGCCGTGTGGACAAATATCGCACAG 3518

Qy   3628 nacaccaaacactaccgttcgcaagatttccgcgcctaccgccaac---aaaccgacctg 3684
Db   3519 GATAAAAGACGCTATGATTCTGATGCGTTCCGTGCTTATCAGCAGCAGAAAACGAACTTA 3578

Qy   3685 cgccaaatcggtatgcagaaaaacctcggcagcgggcgcgtcggcatcctgttttcgcac 3744
Db   3579 CGTCAAATTGGGGTGCAAAAAGCCTTAGCTAATGGACGAATTGGGGCAGTTTTCTCGCAT 3638

Qy   3745 aaccggaccgaaaacancttcgacgacggcatcggcaactcggcacggcttgcccacggc 3804
Db   3639 AGCCGTTCAGATAATACCTTTGATGAACAGGTTAAAAATCACGCGACATTAACGATGATG 3698

Qy   3805 gccgttttcgggcaatacggcatcggcaggttcgacatcggcatcagcacgggcgcgggt 3864
Db   3699 TCGGGTTTTGCCCAATATCAATGGGGCGATTTACAATTTGGTGTAAACGTGGGAACGGGA 3758

Qy   3865 tttagcagcggcantctntcagacggcatcggaggcaaaatccgccgccgcgtgctgcat 3924
Db   3759 ATCAGTGCGAGTAAAATGGCTGAAGAACAAAGCCGAAAAATTCATCGAAAAGCGATAAAT 3818

Qy   3925 tacggcattcaggcacgataccgcgccggtttcggcggattcggcatcgaaccgtacatc 3984
Db   3819 TATGGCGTGAATGCAAGTTATCAGTTCCGTTTAGGGCAATTGGGCATTCAGCCTTATTTT 3878

Qy   3985 ggcgcaacgcgctatttcgtccaaaaagcggattaccgctacgaaaacgtcaatatcgcc 4044
Db   3879 GGAGTTAATCGCTATTTTATTGAACGTGAAAATTATCAATCTGAGGAAGTGAGAGTGAAA 3938

Qy   4045 acccccggtcttgcgttcaaccgntaccgngcgggcattaaggcagattattcattcaaa 4104
Db   3939 ACGCCTAGCCTTGCATTTAATCGCTATAATGCTGGCATTCGAGTTGATTATACATTTACT 3998

Qy   4105 ccggcgcaacacatntccatcacnccttatttnagcctgtcctataccgatgccgcttcg 4164
Db   3999 CCGACAGATAATATCAGCGTTAAGCCTTATTTCTTCGTCAATTATGTTGATGTTTCAAAC 4058

Qy   4165 ggcaaagtccgaacacgcgtcaataccgcngtattggctcaggatttcggcaaaacccgc 4224
Db   4059 GCTAACGTACAAACCACGGTAAATCTCACGGTGTTGCAACAACCATTTGGACGTTATTGG 4118

Qy   4225 agtgcggaatggggcgtaaacgccgaaatcaaaggtttcacgctgtccntccacgctgcc 4284
Db   4119 CAAAAAGAAGTGGGATTAAAGGCAGAAATTTTACATTTCCAAATTTCCGCTTTTATCTCA 4178

Qy   4285 gccgccaaaggnccgcaactggaagcgcaacacagcgcgggcatcaaattaggctaccgc 4344
Db   4179 AAATCTCAAGGTTCACAACTCGGCAAACAGCAAAATGTGGGCGTGAAATTGGGCTATCGT 4238

Qy   4345 tggtaa 4350
Db   4239 TGGTAA 4244


RESULT 14
U32710/c
LOCUS       U32710                 13818 bp    DNA     linear   BCT 29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 25 of 163 of the complete genome.
ACCESSION   U32710 L42023
VERSION     U32710.1  GI:1573200
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
  ORGANISM  Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 13818)
  AUTHORS   Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
REFERENCE   2  (bases 1 to 13818)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
REFERENCE   3  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The H. influenzae sequence has been updated by R. Fleischmann. New
            database matches have been assigned, product names have been
            improved, and a number of frame shifts have been corrected. We
            gratefully acknowledge the work of Tatusov et. al. We have
            incorporated their annotation into the /notes fields of the
            corresponding H. influenzae genes
REFERENCE   5  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
            Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The whole genome was shifted by 588 nucleotides for a new start
COMMENT     On Sep 30, 1996 this sequence version replaced gi:1220911.
FEATURES             Location/Qualifiers
     source          1. .13818
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            complement(134. .343)
                     /gene="HI0235"
     CDS             complement(134. .343)
                     /gene="HI0235"
                     /note="similar to GB:L29458 SP:P36675 PID:473421 percent
                     identity: 73.81; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21905.1"
                     /db_xref="GI:1573201"
                     /translation="MRKKQKSAVENETVYLHTRGVIKDNAVMALLGDKLFRQRVEKKR
                     KGKGSYQRKAKHPGKIFEKPDYKFF"
     gene            complement(411. .761)
                     /gene="HI0236"
     CDS             complement(411. .761)
                     /gene="HI0236"
                     /note="similar to PID:1805555 percent identity: 45.74;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="arsenate reductase, putative"
                     /protein_id="AAC21906.1"
                     /db_xref="GI:1573202"
```

```
                     /translation="MSVIIYHNPHCSKSRETLALLENKGIQPIIELYLQKQYSVNELQ
                     SIAKKLGIDDVRQMMRTKDELYKSLNLDNLDLSQAELFKAISEHSALIERPIVINGDK
                     AKIGRPPETVLEIL"
     gene            840. .1888
                     /gene="HI0237"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     GB:U00096 SP:P77406 PID:1621505 PID:1788838 percent
                     identity: 55.75; identified by sequence similarity;
                     putative"
                     /pseudo
     gene            complement(2003. .2980)
                     /gene="HI0239"
     CDS             complement(2003. .2980)
                     /gene="HI0239"
                     /note="similar to SP:P19674 GB:X56175 PID:581231 GB:U00096
                     PID:1773093 percent identity: 48.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-export membrane protein (secF)"
                     /protein_id="AAC21907.1"
                     /db_xref="GI:1573204"
                     /translation="MMKLFTKDKDGHFIREINGIKLPFPLTEFMKVRKLGYILSALLM
                     VISLFFIITKGFNWGLDFTGGVVFDTHFSQSANLEQIRSKLHENGIESPIVQTTGSVQ
                     DVMIRLPASNNDSTIGEHVKSMLQNVDKDIQIRSIEFVGPNVGEELAQGAVYATLATL
                     AMVLIYVGSRFEWRLGFGSIASLAHDVIITLGVFSALQIEIDLTFVAAILSVVGYSIN
                     DSIVVFDRVRENFRKIRRLDTIDIIDISLTQTLSRTIITSVTTLVVVMALFFFGGPSI
                     HNFSLALLVGIGFGTYSSIFVAIAIAYDVGLRREHMIPPKVDKEIDELP"
     gene            complement(2988. .4838)
                     /gene="HI0240"
     CDS             complement(2988. .4838)
                     /gene="HI0240"
                     /note="similar to SP:P19673 GB:S68715 GB:X56175 PID:581230
                     PID:545175 percent identity: 59.74; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-export membrane protein (secD)"
                     /protein_id="AAC21908.1"
                     /db_xref="GI:1573205"
                     /translation="MLNRYPLWKNLMVIFIVAIGILYSLPNIYGEDPAVQISGTRGQE
                     ANTSVLGQVQDVLKTNNLPTKSIVLENGSILARFTNTDDQLLAKDKIAERLGNNYTTA
                     LNLAPATPAWLSMFGANPMKWGLDLRGGVRFLMEVDMNATLVKRQEQLQDSLRGELRK
                     EKIQYTAIKNTEHFGTLITLANVSQRAKAERIIRQLHPTLDITEPPDADSINLGLSTAA
                     LNEARDLAIEQNLTILRKRVAELGVAEAVIQRQGAERIVIELPGVQDTARAKEILGAT
                     ATLEFRIVNQNVTADAISRNMLPADSEVKYDRQGHPVALFKRAVLGGEHIINSSSGLD
                     QHSSTPQVSVTLDSEGGEIMSQTTKKYYKKPMATLYVEYKDNGKKDENGKTILEKHEE
                     VINVATIQGRFGSNFQITGVDSIAEAHNLSTLLKSGALIAPIQIVEERTIGPSLGAQN
                     VEQGINASLWGLVAVIAFMLFYYKMFGVIASFALVINIVLLVGLMSILPGATLSMPGI
                     AGIVLTLGMSVDANVLIFERIKEEIRNGRSIQQAINEGYNGAFTSIFDANLTTILTAI
                     ILYAVGTGPIQGFAITLSLGVAISMFTAITGTRALVNALYGGKQLKKLLI"
     gene            complement(4908. .5201)
                     /gene="HI0241"
     CDS             complement(4908. .5201)
                     /gene="HI0241"
                     /note="similar to PID:599589 percent identity: 62.37;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21909.1"
                     /db_xref="GI:1573206"
                     /translation="MEAQSPMSTLFIFVIFGLIFYFMIYRPQAKRNKEHKKLMSELAK
                     GTEVLTAGGVIGKITKVTEGSDSIVIALNDTTEITINRNYIVSVLPKGSLKSL"
     gene            complement(5309. .5530)
                     /gene="HI0242"
     CDS             complement(5309. .5530)
                     /gene="HI0242"
                     /note="similar to GB:L77117 SP:Q58397 PID:1591652 percent
                     identity: 35.71; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC21910.1"
                     /db_xref="GI:1573207"
                     /translation="MKYQLNLTALRCPIPLLSAKKALKNLDKNDELMLILNLESAVEN
                     FSIFAEENSVALVEQYYASEKEFIVILKK"
     gene            complement(5527. .6045)
                     /gene="HI0243"
     CDS             complement(5527. .6045)
                     /gene="HI0243"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. influenzae predicted coding region HI0243"
                     /protein_id="AAC21913.1"
                     /db_xref="GI:1573208"
                     /translation="MQILEPQQFATWNEPIEMLYACHSKVKRFCRQLSILPDYLEKHG
                     YTQAVLNDVEQILSYFNRAAPLHHDDEELDFFPQLVKVAPQTQTSIDELEKQHEYLHE
                     NWNALSVQLEELISEQRQDIDKHLIERFIQGYDRHIALEEPLFEMGRECLSADILTEM
                     GKHMSARRQVKE"
     gene            complement(6110. .7258)
                     /gene="HI0244"
     CDS             complement(6110. .7258)
                     /gene="HI0244"
                     /note="similar to GB:M63939 SP:P19675 GB:S68715 GB:X56175
                     PID:147966 percent identity: 81.60; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tRNA-guanine transglycosylase (tgt)"
                     /protein_id="AAC21911.1"
                     /db_xref="GI:1573209"
                     /translation="MKYELDKTSGNARRGRLVFERPQGTFSVETPAFMPVGTYGTVKG
                     MTPEEVRATGAEILLGNTFHLWLRPGQEVMRKHGDLHDFMQWHRPILTDSGGFQVFSL
                     GKLRKITEEGVKFQNPINGERIFLSPEKSMEIQYDLGSDIVMIFDECTPYPATFDYAK
                     KSMEMSLRWAKRSRDRFDELGNKNALFGIIQGGVFEELRKVSLEGLVNIGFDGYAVGG
                     LAVGEPKEDMHRILEYICPQIPADKPRYLMGVGKPEDLVEGVRRGIDMFDCVMPTRNA
                     RNGHLFVTDGIVKIRNAKYRDDTSPLDPECDCYTCKNYTKAYLYHLDKCGEILGARLN
                     TIHNLRYYQRLMAEIRQAIEDDRFDDFVVEFYARMGKPVPPLQLADKS"
     gene            complement(7602. .8693)
                     /gene="HI0245"
     CDS             complement(7602. .8693)
                     /gene="HI0245"
```

```
  Query Match             28.9%;  Score 1255.8;  DB 1;  Length 13818;
  Best Local Similarity   57.7%;  Pred. No. 2e-276;
  Matches 2515;  Conservative    0;  Mismatches 1616;  Indels  225;  Gaps   10;

Qy      73 cctgcttacttagccatatgcctgtcgttcggcattcttccccaagcttgggcgggacac 132
           | |  ||  ||| ||   ||  | || || || ||     | ||||| ||||| || ||
Db   13610 CTTAATTTTTTAACCGCTTGTGTTTCATTAGGGATAGCATCACAAGCCTGGGCAGGTCAT 13551

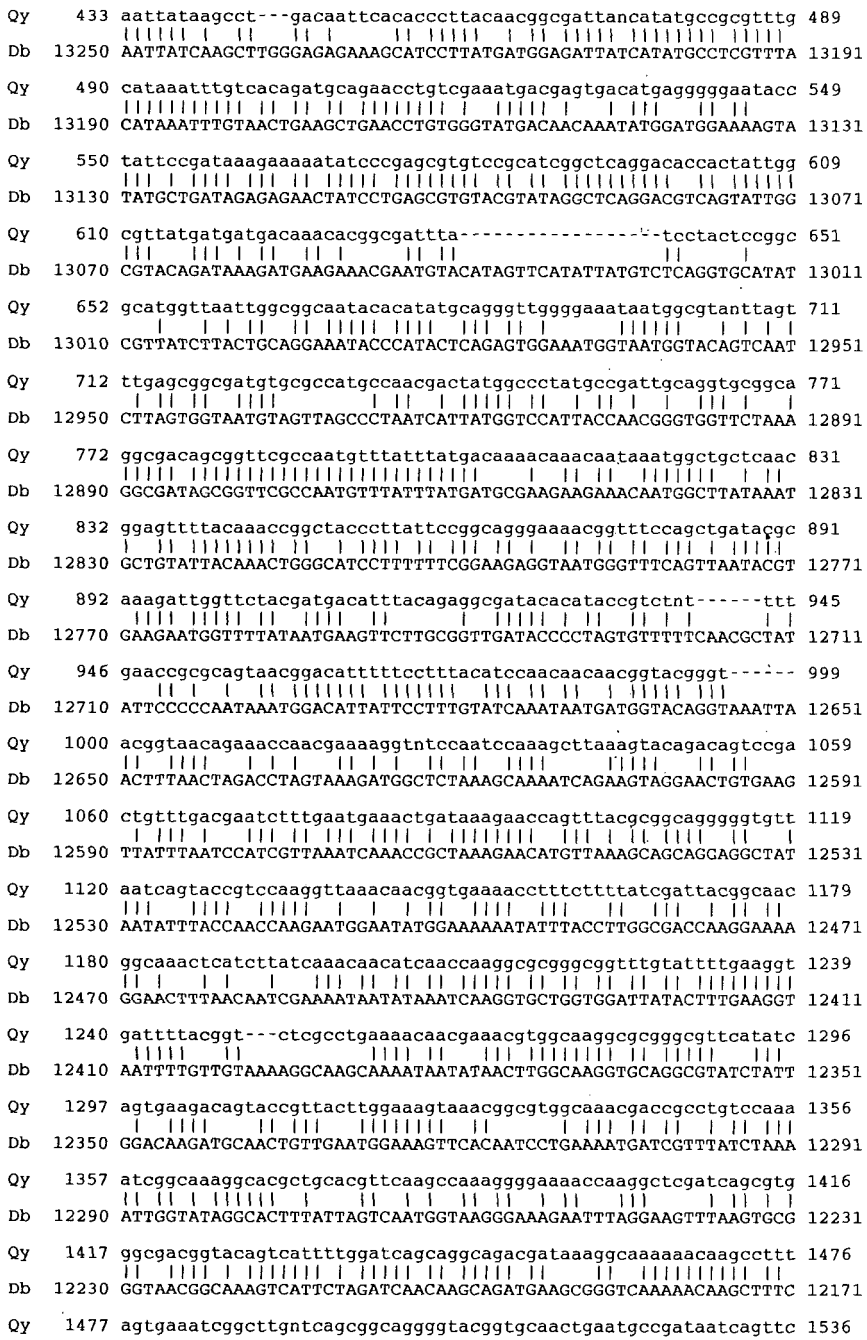
Qy     133 acttatttcggcatcaactaccaatactatcgcgactttgccgaaaataaaggcaagttt 192
           |||||||| || ||  |||||||||| ||||| || |||||||| |||||||| |||||
Db   13550 ACTTATTTTGGGATTGACTACCAATATTATCGTGATTTTGCCGAGAATAAAGGGAAGTTC 13491

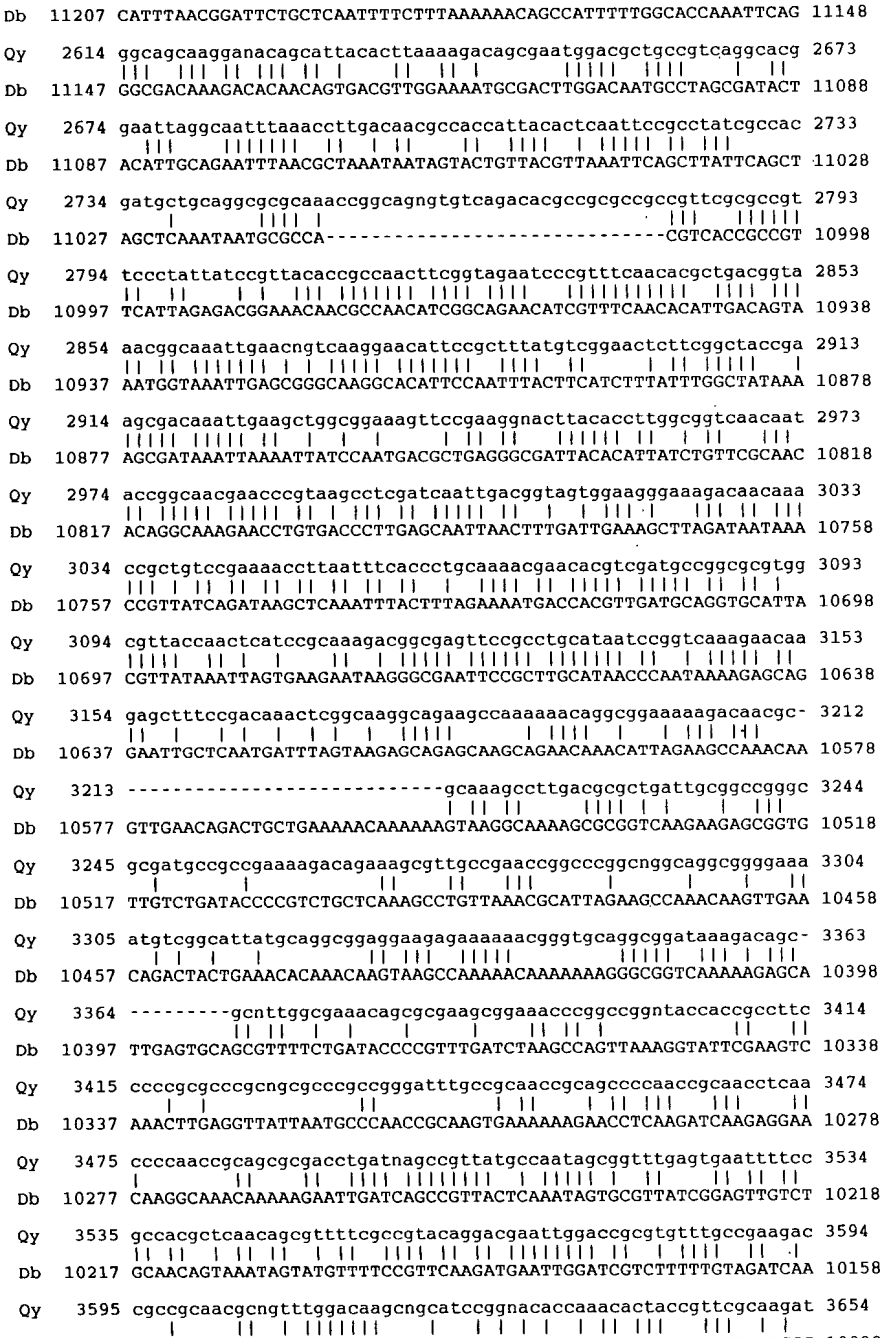
Qy     193 gcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttggtcggcaaatca 252
           |||| ||||| ||| |||||||||| || |||||| ||||| | || || |||| ||||
Db   13490 ACAGTTGGGGCTAAAAATATTGAGGTTTATAACAAAGAAGGGCAATTAGTTGGCACATCA 13431

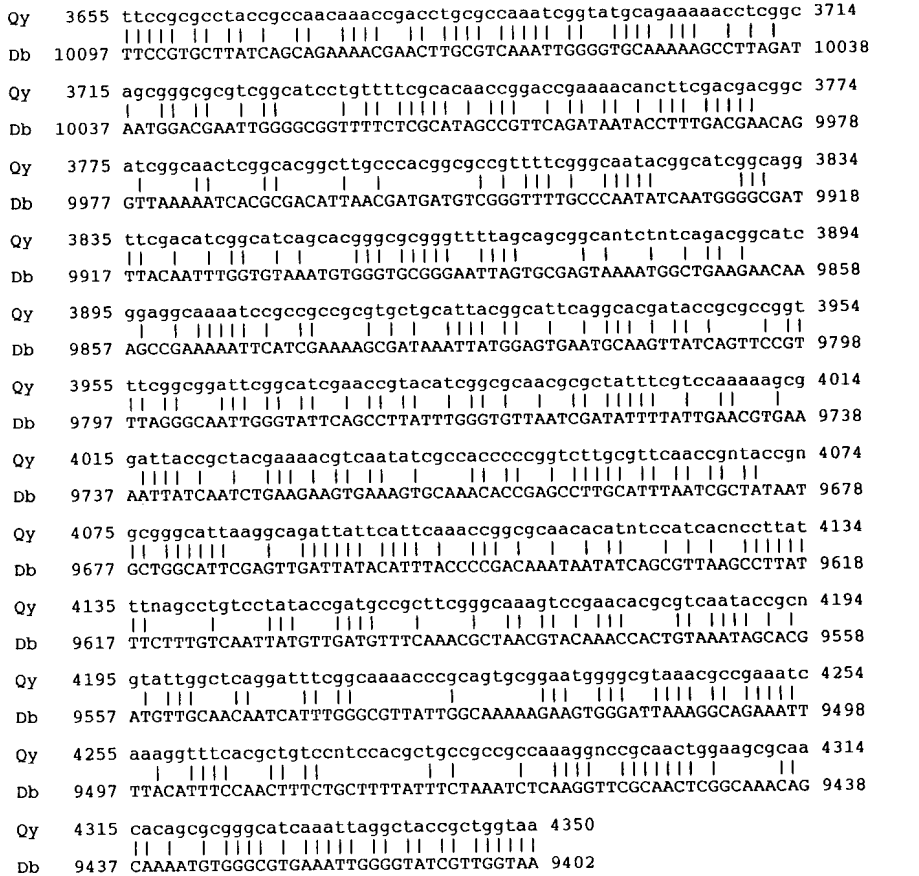
Qy     253 atgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattg 312
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   13430 ATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCATTA 13371

Qy     313 gtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgatttt 372
           || |||||||| |||||||||||||||||||||||||||||||| |||||| | |||||||||
Db   13370 GTAGGCGATCAGTATATTGTGAGCGTGGCACATAACGGCGGATATAACGATGTTGATTTT 13311

Qy     373 ggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaattgtgaaaagaaat 432
           ||||| |||||| | || || |||||||||| ||| |||| |||||||| |||||||||
Db   13310 GGTGCAGAAGGACGAAACCCTGATCAGCACCGCTTTACTTATCAAATTGTAAAAAGAAAT 13251
```


Qy 433 aattataagcct---gacaattcacacccttacaacggcgattancatatgccgcgtttg 489
Db 13250 AATTATCAAGCTTGGGAGAGAAAGCATCCTTATGATGGAGATTATCATATGCCTCGTTTA 13191
Qy 490 cataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgaggggaatacc 549
Db 13190 CATAAATTTGTAACTGAAGCTGAACCTGTGGGTATGACAACAAATATGGATGGAAAAGTA 13131
Qy 550 tattccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacaccactattgg 609
Db 13130 TATGCTGATAGAGAGAACTATCCTGAGCGTGTACGTATAGGCTCAGGACGTCAGTATTGG 13071
Qy 610 cgttatgatgatgacaaacacggcgattta----------tcctactccggc 651
Db 13070 CGTACAGATAAAGATGAAGAAACGAATGTACATAGTTCATATTATGTCTCAGGTGCATAT 13011
Qy 652 gcatggttaattggcggcaatacacatatgcagggttggggaaataatggcgtanttagt 711
Db 13010 CGTTATCTTACTGCAGGAAATACCCATACTCAGAGTGGAAATGGTAATGGTACAGTCAAT 12951
Qy 712 ttgagcggcgatgtgcgccatgccaacgactatggccctatgccgattgcaggtgcggca 771
Db 12950 CTTAGTGGTAATGTAGTTAGCCCTAATCATTATGGTCCATTACCAACGGGTGGTTCTAAA 12891
Qy 772 ggcgacagcggttcgccaatgtttatttatgacaaaacaaacaataaatggctgctcaac 831
Db 12890 GGCGATAGCGGTTCGCCAATGTTTATTTATGATGCGAAGAAGAAACAATGGCTTATAAAT 12831
Qy 832 ggagttttacaaaccggctacccttattccggcagggaaaacggtttccagctgatacgc 891
Db 12830 GCTGTATTACAAACTGGGCATCCTTTTTTCGGAAGAGGTAATGGGTTTCAGTTAATACGT 12771
Qy 892 aaagattggttctacgatgacatttacagaggcgatacacataccgtctnt------ttt 945
Db 12770 GAAGAATGGTTTTATAATGAAGTTCTTGCGGTTGATACCCCTAGTGTTTTTCAACGCTAT 12711
Qy 946 gaaccgcgcagtaacggacatttttcctttacatccaacaacaacggtacgggt------ 999
Db 12710 ATTCCCCCAATAAATGGACATTATTCCTTTGTATCAAATAATGATGGTACAGGTAAATTA 12651
Qy 1000 acggtaacagaaaccaacgaaaaggtntccaatccaaagcttaaagtacagacagtccga 1059
Db 12650 ACTTTAACTAGACCTAGTAAAGATGGCTCTAAAGCAAAATCAGAAGTAGGAACTGTGAAG 12591
Qy 1060 ctgtttgacgaatctttgaatgaaactgataaagaaccagtttacgcggcaggggtgtt 1119
Db 12590 TTATTTAATCCATCGTTAAATCAAACCGCTAAAGAACATGTTAAAGCAGCAGGAGGCTAT 12531
Qy 1120 aatcagtaccgtccaaggttaaacaacggtgaaaacctttctttttatcgattacggcaac 1179
Db 12530 AATATTTACCAACCAAGAATGGAATATGGAAAAAATATTTACCTTGGCGACCAAGGAAAA 12471
Qy 1180 ggcaaactcatcttatcaaacaacatcaaccaaggcgcgggcggtttgtattttgaaggt 1239
Db 12470 GGAACTTTAACAATCGAAAATAATATAAATCAAGGTGCTGGTGGATTATACTTTGAAGGT 12411
Qy 1240 gattttacggt---ctcgcctgaaaacaacgaaacgtggcaaggcgcgggcgttcatatc 1296
Db 12410 AATTTTGTTGTAAAAGGCAAGCAAAATAATATAACTTGGCAAGGTGCAGGCGTATCTATT 12351
Qy 1297 agtgaagacagtaccgttacttggaaagtaaacggcgtggcaaacgaccgcctgtccaaa 1356
Db 12350 GGACAAGATGCAACTGTTGAATGGAAAGTTCACAATCCTGAAAATGATCGTTTATCTAAA 12291
Qy 1357 atcggcaaaggcacgctgcacgttcaagccaaaggggaaaaccaaggctcgatcagcgtg 1416
Db 12290 ATTGGTATAGGCACTTTATTAGTCAATGGTAAGGGAAAGAATTTAGGAAGTTTAAGTGCG 12231
Qy 1417 ggcgacggtacagtcattttggatcagcaggcagacgataaaggcaaaaaacaagccttt 1476
Db 12230 GGTAACGGCAAAGTCATTCTAGATCAACAAGCAGATGAAGCGGGTCAAAAACAAGCTTTC 12171
Qy 1477 agtgaaatcggcttgntcagcggcaggggtacggtgcaactgaatgccgataatcagttc 1536
Db 12170 AAAGAAGTTGGCATTGTAAGCGGTCGAGCAACAGTTCAATTAAATAGTACAGATCAAGTT 12111
Qy 1537 aaccccgacaaactctatttcggctttcgcggcggacgtttggatttaaacgggcattcg 1596
Db 12110 GATCCTAACAATATCTATTTCGGATTTCGTGGTGGTCGCTTAGATCTTAACGGGCATTCA 12051
Qy 1597 ctttcgttccaccgtattcaaaataccgatgaaggggcgatgattgncnatcataatgcc 1656
Db 12050 TTAACCTTTAAGCGTATCCAAAATACGGACGAGGGCGCGATGATTGTGAACCATAATACA 11991
Qy 1657 acaacaacatccaccgttaccattacagggaatgaaagtatta---cacaaccgagtggt 1713
Db 11990 ACTCAAGTCGCTAATATTACTATTACTGGGAACGAAAGTATTACTGCTCCATCTAATAAA 11931
Qy 1714 aagaatatcaatagacttaattacagcaaagaaattgcctacaacggttggtttggcgag 1773
Db 11930 AAGAATATTAATAAACTTGATTACAGCAAAGAAATTGCCTACAACGGTTGGTTTGGCGAA 11871
Qy 1774 aaagatacgaccaaaacgaacgggcggctcaaccttgtttaccagcccgccgcagaagac 1833
Db 11870 ACAGATAAAAATAAACACAATGGACGATTAAACCTTATTTATAAACCAACCACAGAAGAT 11811
Qy 1834 cgcacccngctgctttccggcggaacaaatttaaacggcaacatcacgcaaacaaacggc 1893
Db 11810 CGTACTTTGCTACTTTCAGGTGGCACAAACTTAAAAGGCGATATTACCCAAACAAAAGGT 11751
Qy 1894 aaactgtttttcagcggcagaccgacaccgcacgcctacaatcatttaggaagcgggtgg 1953
Db 11750 AAACTATTTTTCAGCGGTAGACCGACACCCCACGCCTACAATCATTTAGACAAACGTTGG 11691
Qy 1954 tcaaaaatggaaggtatcccacaaggagaaatcgtgtgggacaacgactggatcnaccgc 2013
Db 11690 TCAGAAATGGAAGGTATCCCACAAGGCGAAATTGTGTGGGATTACGATTGGATCAACCGT 11631
Qy 2014 acgtttaaagcggaaaatttccatattcagggcgggcaggcggtgatttcccgcaatgtt 2073
Db 11630 ACATTTAAAGCTGAAAACTTCCAAATTAAAGGCGGAAGTGCGGTGGTTTCTCGCAATGTT 11571
Qy 2074 gccaaagtggaaggcgattgncatttgagcaatcacgcccaagcagtttttggtgtcgca 2133
Db 11570 TCTTCAATTGAGGGAAATTGGACAGTCAGCAATAATGCAAATGCCACATTTGGTGTTGTG 11511
Qy 2134 ccgcatcaaagccatacaatctgtacacgttcggactggacnggtctgacaaattgtgtc 2193
Db 11510 CCAAATTAACAAAATACCATTTGCACGCGTTCAGATTGGACAGGATTAACGACTTGTAAA 11451
Qy 2194 gaanaaancattaccgacgataaagtgattgcttcattgactaagacngacntnagcggc 2253
Db 11450 ACAGTTAATTTAACCGATACAAAAGTTATTAATTCCATACCCATAACACAAATTAATGGC 11391
Qy 2254 antgtnagnctnnccnatnacgntnnttnaaanctcncnggcntgcnncactnaanggc 2313
Db 11390 TCTATTAATTTAACTAATAATGCAACAGTGAATATTCATGGTTTAGCAAAACTTAATGGT 11331
Qy 2314 aatcttagtgcaaatggcgatacacgttatacagtcagccacaacgccacccaaaacggc 2373
Db 11330 AATGTCACTTTAATAGATCACAGCCAATTTACATTGAGCAACAATGCCACCCAAACAGGC 11271
Qy 2374 aaccttagcctcgtgggcaatgcccaagcaacatttaatcaagccacattaaacggcaac 2433
Db 11270 AA---------- 11267
Qy 2434 ncatcggnttcgggcaatgcttcatttaatctaagcaacaacgccgcacaaaacggcagt 2493
Db 11268 ----------T 11268
Qy 2494 ctgacgctttccgacaacgctaaggcaaacgtaagccattccgcactcaacggcaatgtc 2553
Db 11267 ATCAAACTTTCAAATCACGCAAATGCAACGGTAAATAATGCCACGTTAAACGGCAATGTG 11208
Qy 2554 tccctagccgataaggcagtattccattttgaaaacagccgctttaccggacaactcagc 2613





RESULT 15
NLIGA3327

LOCUS	NLIGA3327 578 bp DNA linear BCT 21-MAY-1998
DEFINITION	Niesseria lactamica IgA protease homologue gene, isolate NL3327.
ACCESSION	AJ001740
VERSION	AJ001740.1 GI:3152399
KEYWORDS	IgA protease.
SOURCE	Neisseria lactamica.
ORGANISM	Neisseria lactamica Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE	1 (bases 1 to 578)
AUTHORS	Sayers,J.R.
TITLE	Direct Submission
JOURNAL	Submitted (20-OCT-1997) Sayers J.R., Division of Molecular Medicine, University of Sheffield, Royal Hallamshire Hospital, Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE	2 (bases 1 to 578)
AUTHORS	Sayers,J.R., Vitowski,S. and Read,R.
TITLE	Strains of Neisseria lactamica carry an IgA1 protease gene homologue and some express IgA1 protease activity
JOURNAL	Unpublished
FEATURES	Location/Qualifiers

```
     source          1. .578
                     /organism="Neisseria lactamica"
                     /isolate="NL3327"
                     /db_xref="taxon:486"
     CDS             <1. .>576
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA04965.1"
                     /db_xref="GI:3152400"
                     /db_xref="SPTREMBL:O69746"
                     /translation="GKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALV
                     GDQYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRL
                     HKFVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDKHGDLSYSGAWLI
                     GGNTHMQGWGNNGVVSLSGDVRHANDYGPMPI"
     repeat_region   88. .95
                     /rpt_type=INVERTED
     repeat_region   139. .148
BASE COUNT      163 a    125 c    156 g    134 t
ORIGIN


  Query Match             12.6%;  Score 550;  DB 1;  Length 578;
  Best Local Similarity   96.7%;  Pred. No. 4.5e-115;
  Matches  559;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy    184 ggcaagtttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttggtc 243
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db      1 GGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 60

Qy    244 ggcaaatcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtg 303
          ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db     61 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTG 120

Qy    304 gcggcattggtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaac 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAAT 180

Qy    364 gttgattttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaattgtg 423
          ||||||||||||| || ||||| |||||||||||||||| |||||||||||||||||||
Db    181 GTTGATTTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG 240

Qy    424 aaaagaaataattataagcctgacaattcacacccttacaacggcgattancatatgccg 483
          |||||||||||| || || ||||||||||||||||||||||||||||||| |||||||||
Db    241 AAAAGAAATAACTACAAACCTGACAATTCACACCCTTACAACGGCGATTACCATATGCCG 300

Qy    484 cgtttgcataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgaggggg 543
          ||  ||||||| || || || |||||||||||||||||||||||| |||||||||||||
Db    301 CGCCTGCATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGG 360

Qy    544 aatacctattccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacaccac 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AATACCTATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCAC 420

Qy    604 tattggcgttatgatgatgacaaacacggcgatttatcctactccggcgcatggttaatt 663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TATTGGCGTTATGATGATGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTAATT 480

Qy    664 ggcggcaatacacatatgcagggttggggaaataatggcgtanttagtttgagcggcgat 723
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    481 GGCGGCAATACACATATGCAGGGTTGGGGAAATAATGGCGTAGTTAGTTTGAGCGGCGAT 540

Qy    724 gtgcgccatgccaacgactatggccctatgccgattgc 761
          |||||||||||||||||||||||||||||||||||||
Db    541 GTGCGCCATGCCAACGACTATGGCCCTATGCCGATTGC 578


RESULT  16
NLIGA3354
LOCUS       NLIGA3354               584 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3354.
ACCESSION   AJ001739
VERSION     AJ001739.1  GI:3152397
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 584)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Neisseria lactamica"
                     /isolate="NL3354"
                     /db_xref="taxon:486"
     CDS             <1. .>582
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04964.1"
                     /db_xref="GI:3152398"
                     /db_xref="SPTREMBL:O69747"
                     /translation="FAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGD
                     QYIVSVAHNGGYNNVDFGAEGSNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLHK
                     FVTDAEPVEMTGDMRGNTYSDKEKYPERVRIGSGHHYWRYDDDDKHGDLSYSGAWLIG
                     GNTHMQGWGNNGVVSLSGDVRHANDYGPMPIAGA"
     repeat_region   82. .88
                     /rpt_type=INVERTED
     repeat_region   134. .142
BASE COUNT      163 a    127 c    159 g    135 t
ORIGIN


  Query Match             12.4%;  Score 540;  DB 1;  Length 584;
  Best Local Similarity   96.2%;  Pred. No. 8.7e-113;
  Matches  562;  Conservative    0;  Mismatches   19;  Indels    3;  Gaps    1;

Qy    190 tttgcagtcggggcgaaagatattgaggtntacaacaaaaaaggggagttggtcggcaaa 249
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      1 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAA 60

Qy    250 tcaatgacaaaagccccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggca 309
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db     61 TCAATGACAAAAGCCCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCA 120

Qy    310 ttggtgggcgatcaatatattgtgagcgtggcacataacggcggctataacaacgttgat 369
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    121 TTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGAT 180

Qy    370 tttggtgcggaaggaagnaatcccgatcagcaccgtttttcttaccaaattgtgaaaaga 429
          ||||||||| || ||||| |||||||||||||||| |||||||||||||||||||||||||
Db    181 TTTGGTGCAGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGA 240

Qy    430 aataattataagcctgacaattcacacccttacaacggcgattancatatgccgcgtttg 489
          ||||| || || ||||||||||||||||||||||||||||||| ||||||||||  ||
Db    241 AATAACTACAAACCTGACAATTCACACCCTTACAACGGCGATTACCATATGCCGCGCCTG 300

Qy    490 cataaatttgtcacagatgcagaacctgtcgaaatgacgagtgacatgaggggaatacc 549
          ||||| || || || |||||||||||||||||||||| |||||||||||||||||||||
Db    301 CATAAGTTCGTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGGAATACC 360

Qy    550 tattccgataaagaaaaatatcccgagcgtgtccgcatcggctcaggacaccactattgg 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TATTCCGATAAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTATTGG 420

Qy    610 cgttatgatgat---gacaaacacggcgatttatcctactccggcgcatggttaattggc 666
          ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    421 CGTTATGATGATGACGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTAATTGGC 480

Qy    667 ggcaatacacatatgcagggttggggaaataatggcgtanttagtttgagcggcgatgtg 726
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    481 GGCAATACACATATGCAGGGTTGGGGAAATAATGGCGTAGTTAGTTTGAGCGGCGATGTG 540

Qy    727 cgccatgccaacgactatggccctatgccgattgcaggtgcggc 770
          |||||||||||||||||||||||||||||||||||||||||||
Db    541 CGCCATGCCAACGACTATGGCCCTATGCCGATTGCAGGTGCGGC 584


RESULT 17
NLIGA823
LOCUS       NLIGA823                572 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL823.
ACCESSION   AJ001737
VERSION     AJ001737.1  GI:3152389
KEYWORDS    IgA protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 572)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Neisseria lactamica"
                     /isolate="NL823"
                     /db_xref="taxon:486"
     CDS             <1. .304
                     /codon_start=2
                     /transl_table=11
                     /product="IgA protease homologue"
                     /protein_id="CAA04962.1"
                     /db_xref="GI:3152390"
                     /db_xref="SPTREMBL:O69748"
                     /translation="KDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGDQYIVS
                     VAHNGGYNNVDFGAEGGNPDQHRFSYQIVKRNNYKPDNSHPYNGDYHMPRLPGISS"
     repeat_region   44. .51
                     /rpt_type=INVERTED
     repeat_region   68. .75
BASE COUNT      158 a    124 c    157 g    133 t
ORIGIN


  Query Match             11.9%;  Score 517.6;  DB 1;  Length 572;
  Best Local Similarity   95.1%;  Pred. No. 1.2e-107;
  Matches  544;  Conservative    0;  Mismatches   23;  Indels    5;  Gaps    1;

Qy    204 gaaagatattgaggtntacaacaaaaaaggggagttggtcggcaaatcaatgacaaaagc 263
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 60

Qy    264 cccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattggtgggcgatca 323
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     61 CCCGATGATTGATTTTTCCGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCA 120

Qy    324 atatattgtgagcgtggcacataacggcggctataacaacgttgattttggtgcggaagg 383
          |||||||||||||||||| |||||||||||||||||| |||||||||||||||| ||
Db    121 ATATATTGTGAGCGTGGCGCATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGG 180

Qy    384 aagnaatcccgatcagcaccgtttttcttaccaaattgtgaaaagaaataattataagcc 443
          | | ||||||||||||||||| |||||||||||||||| |||||||||| || || ||
Db    181 AGGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTAAAAAGAAATAACTACAAACC 240

Qy    444 tgacaattcacacccttacaacggcgattancatatgccgcg-----tttgcataaattt 498
          |||||||||||||||||||||||||||||||| |||||||||||      | |||||| ||
Db    241 TGACAATTCACACCCTTACAACGGCGATTACCATATGCCGCGCCTGCCTGGCATAAGTTC 300

Qy    499 gtcacagatgcagaacctgtcgaaatgacgagtgacatgaggggggaatacctattccgat 558
          || || |||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    301 GTAACCGATGCAGAACCTGTCGAAATGACGGGTGACATGAGGGGGAATACCTATTCCGAT 360

Qy    559 aaagaaaaatatcccgagcgtgtccgcatcggctcaggacaccactattggcgttatgat 618
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGAAAAATATCCCGAGCGTGTCCGCATCGGCTCAGGACACCACTATTGGCGTTATGAT 420

Qy    619 gatgacaaacacggcgatttatcctactccggcgcatggttaattggcggcaatacacat 678
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    421 GATGACAAACACGGCGATTTATCCTACTCCGGCGCATGGTTGATTGGCGGCAATACACAT 480

Qy    679 atgcagggttggggaaataatggcgtanttagtttgagcggcgatgtgcgccatgccaac 738
          |||||||||||||||||||||||||| |||||||||||| ||||||||||||||||||
Db    481 ATGCAGGGTTGGGGAAATAATGGCGTAGTTAGTTTGAGCGGTGATGTGCGCCATGCCAAC 540

Qy    739 gactatggccctatgccgattgcaggtgcggc 770
          |||||||||||||||||||||||||||||||
Db    541 GACTATGGCCCTATGCCGATTGCAGGTGCGGC 572


RESULT 18
NLIGA3293
LOCUS       NLIGA3293               572 bp    DNA     linear   BCT 21-MAY-1998
DEFINITION  Niesseria lactamica IgA protease homologue gene, isolate NL3293.
ACCESSION   AJ001738
VERSION     AJ001738.1  GI:3152385
KEYWORDS    IgA1 protease.
SOURCE      Neisseria lactamica.
  ORGANISM  Neisseria lactamica
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1997) Sayers J.R., Division of Molecular
            Medicine, University of Sheffield, Royal Hallamshire Hospital,
            Sheffield, S. Yorks, S10 2JF, ENGLAND
REFERENCE   2  (bases 1 to 572)
  AUTHORS   Sayers,J.R., Vitowski,S. and Read,R.
  TITLE     Strains of Neisseria lactamica carry an IgA1 protease gene
            homologue and some express IgA1 protease activity
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Neisseria lactamica"
                     /isolate="NL3293"
                     /db_xref="taxon:486"
     CDS             <1. .>570
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA04963.1"
                     /db_xref="GI:3152386"
                     /db_xref="SPTREMBL:O69745"
                     /translation="KDIEVYNKKGELVGKSMTKAPMIDFSVVSRNGVAALVGDQYIVS
                     VAHNVGYRDVDFGAEGSNPDQHRFSYKIAKRNNYKNDETHPYEKDYHNPRLHKFVTEA
                     APIDMTSDMDGNKYTDRTKYPERVRIGSGWQFWRNDQDKGDQVAGAYHYLTAGNTHNQ
                     GGAGGGWSSLSGDVRHAGNYGPIPIAAGSS"
     repeat_region   67. .74
                     /rpt_type=INVERTED
     repeat_region   118. .127
BASE COUNT      169 a    119 c    154 g    130 t
ORIGIN


  Query Match             7.7%;  Score 332.8;  DB 1;  Length 572;
  Best Local Similarity   73.6%;  Pred. No. 2.3e-65;
```

```
  Matches  421;  Conservative    0;  Mismatches  151;  Indels    0;  Gaps    0;

Qy    205 aaagatattgaggtntacaacaaaaaaggggagttggtcggcaaatcaatgacaaaagcc 264
          ||||||||||||||| ||||||||||||||||||||||||||| |||||||||||||||||
Db      1 AAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGAAAATCAATGACAAAAGCC 60

Qy    265 ccgatgattgatttttctgtggtgtcgcgtaacggcgtggcggcattggtgggcgatcaa 324
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     61 CCGATGATTGATTTTTCTGTGGTATCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAA 120

Qy    325 tatattgtgagcgtggcacataacggcggctataacaacgttgattttggtgcggaagga 384
          |||||||||||||||| ||||| |  || |||  | | ||||||||||||||||| |||
Db    121 TATATTGTGAGCGTGGCGCATAATGTAGGTTATCGCGATGTTGATTTTGGTGCGGAGGGA 180

Qy    385 agnaatcccgatcagcaccgtttttcttaccaaattgtgaaaagaaataattataagcct 444
          || |||||||||||||||||||||||||  | ||||  ||| | ||||||||||| ||
Db    181 AGCAATCCCGATCAGCACCGTTTTTCTTATAAGATTGCAAAACGGAATAATTATAAAAAC 240

Qy    445 gacaattcacacccttacaacggcgattancatatgccgcgtttgcataaatttgtcaca 504
          ||  |  | || |||||  |    ||||| || |  || || || |||||||||| ||
Db    241 GATGAAACGCATCCTTATGAGAAAGATTACCACAACCCACGCTTACATAAATTTGTTACG 300

Qy    505 gatgcagaacctgtcgaaatgacgagtgacatgaggggaatacctattccgataaagaa 564
          || || |  ||  | || |||||   ||| |||   || || |  ||| | ||||
Db    301 GAAGCCGCCCCGATTGATATGACTTCTGATATGGACGGCAACAAATATACAGATAGGACG 360

Qy    565 aaatatcccgagcgtgtccgcatcggctcaggacaccactattggcgttatgatgatgac 624
          |||||||||| || ||||| |||||||| ||    || | |||||  | ||| | |||
Db    361 AAATATCCCGAACGCGTCCGTATCGGCTCCGGGTGGCAGTTTTGGCGAAACGATCAAGAC 420

Qy    625 aaacacggcgatttatcctactccggcgcatggttaattggcggcaatacacatatgcag 684
          |||  ||   |  |  ||  | |       | | |  |||||||||| |  ||
Db    421 AAAGGCGATCAAGTTGCCGGCGCATATCATTACCTGACAGCAGGCAATACACACAATCAA 480

Qy    685 ggttggggaaataatggcgtanttagtttgagcggcgatgtgcgccatgccaacgactat 744
          ||  | | |       |||     |||||||||||||||||||||||||   | ||||
Db    481 GGCGGAGCAGGGGGCGGCTGGTCAAGTTTGAGCGGCGATGTGCGCCATGCAGGCAACTAC 540

Qy    745 ggccctatgccgattgcaggtgcggcaggcga 776
          || || || || ||||| |  |   || || |
Db    541 GGTCCCATTCCTATTGCGGCGGGCTCAAGCAA 572


RESULT  19
NMIGA476
LOCUS       NMIGA476                1492 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NG44/76).
ACCESSION   X82481
VERSION     X82481.1  GI:732863
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1492)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1492)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1492
                     /organism="Neisseria meningitidis"
                     /strain="NG44/76"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57864.1"
                     /db_xref="GI:732864"
                     /db_xref="SPTREMBL:Q51165"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHH
                     WNITFGTNSKIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDY
                     TVKGINNDITWLGAGIDVTDGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKV
                     GDGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANG
                     NDLTFEHIRNVDEGARIVNHNTSHASTITLTGKSLITNPNSLSVHS"
BASE COUNT      500 a    317 c    322 g    353 t
ORIGIN


  Query Match             4.5%;  Score 197.6;  DB 1;  Length 1492;
  Best Local Similarity   61.0%;  Pred. No. 2.2e-34;
  Matches  336;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | |||||||||   |  | ||     | ||||||||||||||||||||
Db    918 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 977

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| |||| ||  ||| |  ||| || ||    || ||||
Db    978 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1037

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || | |  || | ||  |  ||   |||| |||| ||      |
Db   1038 AGGTGCGGGGATTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1097

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1098 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1157

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || |||| || || || |   |
Db   1158 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1217

Qy   1458 aggcaaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1218 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1277

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1278 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1337

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaatacgatgaaggggcgat 1637
          ||||   || ||  ||    | || | | || |  ||   |||||||| |||
Db   1338 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1397

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
          |||| | | || ||  | |       | |||||  | ||  | || || ||       ||
Db   1398 CATTGTCAACCACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTAAT 1457

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1458 TACCAATCCGA 1468


RESULT  20
AF012208
LOCUS       AF012208                2983 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4400 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012208
```

```
VERSION     AF012208.1  GI:2529499
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2983)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2983)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2983
                     /organism="Neisseria meningitidis"
                     /strain="Z4400"
                     /db_xref="taxon:487"
                     /note="serogroup C, ET-37 complex; isolated in Mali, 1990"
     mat_peptide     1. .>2983
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2983
                     /gene="iga"
                     /allele="iga5"
     CDS             <1. .>2983
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45791.2"
                     /db_xref="GI:12597968"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHHWNITFGTNSKIGSTAVRLAGNE
                     KDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGINNDITWLGAGIDVTDGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     SHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     RKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPVPSPATNTASQAQT
                     DSAQIAKPQNIVVAPPSPQAN"
BASE COUNT      978 a    679 c    639 g    687 t
ORIGIN
```

```
  Query Match             4.5%;  Score 197.6;  DB 1;  Length 2983;
  Best Local Similarity   61.0%;  Pred. No. 2.4e-34;
  Matches  336;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db   1028 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1087

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| |||| || |||     |  ||| || ||    || ||||
Db   1088 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1147

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || | |  || | ||  |  ||   |||| |||| ||      |
Db   1148 AGGTGCGGGGATTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1207

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
          |||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1208 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1267

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || ||||||||||| || ||| || |||| || || || |   |
Db   1268 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1327

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1328 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1387

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1388 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1447

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | || | |  || | ||  ||    |||||||| |||
Db   1448 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1507

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
          |||| | | || ||  | |      | |||||  | ||  | || || ||       ||
Db   1508 CATTGTCAACCACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTAAT 1567

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1568 TACCAATCCGA 1578
```

```
RESULT  21
AF012209
LOCUS       AF012209                2995 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4081 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012209
VERSION     AF012209.1  GI:2529501
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Neisseria meningitidis"
                     /strain="Z4081"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup V; isolated in China, 1979;
                     also present in strain Z4063, serogroup A, subgroup VII,
                     isolated in China, 1979"
     mat_peptide     1. .>2995
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2995
                     /gene="iga"
                     /allele="iga9"
     CDS             <1. .>2995
                     /gene="iga"
```

```
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45792.2"
                     /db_xref="GI:12597969"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADKIKQRDNAGTIKGNGEHHWNITFGTNSKIGSTAVRLAGNE
                     KDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGINNDITWLGAGIDVTDGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     SHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     RKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANQYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQK
                     ATQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      984 a    683 c    642 g    686 t
ORIGIN


  Query Match             4.5%;  Score 197.6;  DB 1;  Length 2995;
  Best Local Similarity   61.0%;  Pred. No. 2.4e-34;
  Matches  336;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy    1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
           | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||| 
Db    1028 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1087

Qy    1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
           |||| ||| ||| |||| |||| || |||     |  ||| || ||    || ||||
Db    1088 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1147

Qy    1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
           ||| |||||  || || | |  || | ||  |  ||   |||| |||| ||      |
Db    1148 AGGTGCGGGGATTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1207

Qy    1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
              |||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db    1208 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1267

Qy    1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
           ||||||     | |  || |||||||||||| || ||| || |||| || || || |   |
Db    1268 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1327

Qy    1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
              |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db    1328 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1387

Qy    1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
            |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db    1388 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1447

Qy    1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db    1448 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1507

Qy    1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
            |||| | | || ||  | |      | |||||  | ||  | || || || ||       ||
Db    1508 CATTGTCAACCACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTAAT 1567

Qy    1698 tacacaaccga 1708
           |||  | ||||
Db    1568 TACCAATCCGA 1578


RESULT  22
```

```
AF235032
LOCUS       AF235032               4659 bp    DNA     linear   BCT 02-MAR-2001
DEFINITION  Neisseria meningitidis IgA1 protease gene, complete cds.
ACCESSION   AF235032
VERSION     AF235032.1  GI:13183004
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4659)
  AUTHORS   Vitovski,S. and Sayers,J.R.
  TITLE     Degenerate specificity of Neisseria meningitidis IgA1 protease
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4659)
  AUTHORS   Vitovski,S. and Sayers,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Division MGM, University of Sheffield, F
            floor, Royal Hallamshire Hospital, Sheffield S10 2JF, UK
FEATURES             Location/Qualifiers
     source          1. .4659
                     /organism="Neisseria meningitidis"
                     /strain="NMB"
                     /db_xref="taxon:487"
     CDS             1. .4659
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="AAK15023.1"
                     /db_xref="GI:13183005"
                     /translation="MKTKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHV
                     KGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFV
                     TEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGT
                     PYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDK
                     QEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADKIRQRDNAGTIKGNGEHHWNITFG
                     TNSKIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGIN
                     NDITWLGAGIDVTDGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVI
                     LNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFE
                     HIRNVDEGARIVNHNTSHASTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRR
                     PIPQGKDLYYKNYRYYALKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKN
                     NQRISGFSGFFGEENGKGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLS
                     GRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQSASFSSGRNV
                     SNITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGN
                     VNLNQNAALVLGKAALWGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNN
                     ASDAQSANQYHTLKINHLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGE
                     PNQEGLNLFDASSVQDRSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRV
                     KPVPSPATNTASQAQTDSAQIAKPQNIVVAPPSPQANQAEEAKRQQAKAEQVKRQQAE
                     AERKSAELAKQKAEAEREARELATRQKAEQERSSAELARRHEKEREAAELSAKQKVEA
                     EREAQALAVRRKAEAEEAKRQAAELARQQEEARKAAELAAKQKAETERKAAEIAEQKA
                     EAEREAAELAKQKAEEEGRQAAQSQPKRRNRRAIPPELSSDATTRALPRIARNSNPDA
                     SDYEKIPLDALEDEDVSESVDTSDKQPQDNTELHEKVETVSLQPRAAQPRAQAAAQAD
                     AVSTNTNSALSDAMASTQSILLDTGASLTRHIAQKSRADAEKNSVWMSNIGYGRDYAS
                     AQYRRFSSKRTQTQIGIDRSLSENMQIGGVLTYSDSQHTFDQASGKNTFVQANLYGKY
                     YLNDAWYVAGDIGAGSLRSRLQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIR
                     YSRLSSADYKLGNDSVKVSSMSVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKG
                     GVNVGGNSFAYKADNQQQYSAGAALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF
                     "
BASE COUNT     1542 a   1074 c   1041 g   1002 t
ORIGIN


  Query Match             4.5%;  Score 197.6;  DB 1;  Length 4659;
  Best Local Similarity   61.0%;  Pred. No. 2.6e-34;
  Matches  336;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy    1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
           | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||| 
Db    1074 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1133

Qy    1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
           |||| ||| |||  |||| ||||  || |||     |  ||| || ||    || ||||
Db    1134 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1193
```

```
Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || | |  || | ||  |  ||   |||| |||| ||      |
Db   1194 AGGTGCGGGGATTGATGTTACCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1253

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
            |||| |  || | ||||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1254 CGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1313

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || |||| || || || |   |
Db   1314 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1373

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||      || ||||| |  |||||||  | |||||  ||  | |
Db   1374 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1433

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
            |||       |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1434 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1493

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | || | |  || |  ||    |||||||| |||
Db   1494 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1553

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
            |||| | | || ||  | |       | |||||  | ||  | || || ||      ||
Db   1554 CATTGTCAACCACAACACAAGCCACGCCTCCACGATCACGCTAACGGGTAAATCTTTAAT 1613

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1614 TACCAATCCGA 1624


RESULT 23
AF012210
LOCUS       AF012210               2995 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4099 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012210
VERSION     AF012210.1  GI:2529503
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Neisseria meningitidis"
                     /strain="Z4099"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup VIII; isolated in China,
                     1975"
     mat_peptide     <1. .>2995
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2995
                     /gene="iga"
                     /allele="iga8"
     CDS             <1. .>2995
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45793.2"
                     /db_xref="GI:12597970"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQHDNAGTIKGYREHHWNITSGTNSKIGSTAVRLAGNE
                     RNANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGNYTVKGANNDITWLGAGIDVADGK
                     KVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQ
                     VGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNT
                     DHTSTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYY
                     ALKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENG
                     KGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSA
                     QKDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVSNITANITATDNAKVN
                     LGYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFDATQIHGNVNLNQNAALVLGKAAL
                     WGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKIN
                     HLSGNGHFHYLTHLAENLGDKVLVKESASGHYQLHVQDKTGEPNQERLNLFDASSVRD
                     RSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRQVKPTHSPATNTASQAQK
                     ATQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      996 a    690 c    632 g    677 t
ORIGIN


  Query Match             4.5%;  Score 196;  DB 1;  Length 2995;
  Best Local Similarity   60.8%;  Pred. No. 5.7e-34;
  Matches  335;  Conservative    0;  Mismatches  213;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | |||||||||   |  | ||     | || ||||||||||||||||||
Db   1028 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAATATCAACCAAGGCGCGGG 1087

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  ||||  |||  || |||      || ||| || ||    || ||||
Db   1088 CGGTCTGTTTTTCAAAGGCAATTACACAGTCAAAGGTGCAAATAATGACATCACCTGGCT 1147

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          |||||||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1148 AGGCGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1207

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
            |||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1208 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1267

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || |||| |  || ||    ||
Db   1268 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAA 1327

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
            |||   || || ||      || ||||| |  |||||||  | |||||  ||  ||
Db   1328 CCAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTGCT 1387

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
            |||       |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1388 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1447

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | || | |  || |  ||    |||||||| |||
Db   1448 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1507

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
            |||| | | || |||  |       || |||||  | ||  | || || ||      ||
Db   1508 CATTGTCAACCACAATACAGACCACACCTCCACAATCACGCTAACGGGTAAATCTTTAAT 1567

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1568 TACCAATCCGA 1578
```

```
RESULT  24
NMIGA771
LOCUS       NMIGA771                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (EN3771).
ACCESSION   X82468
VERSION     X82468.1  GI:732865
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="EN3771"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57851.1"
                     /db_xref="GI:732866"
                     /db_xref="SPTREMBL:Q57309"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYT
                     VKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI"
BASE COUNT      492 a    321 c    325 g    351 t
ORIGIN


  Query Match             4.4%;  Score 192.8;  DB 1;  Length 1489;
  Best Local Similarity   60.8%;  Pred. No. 2.8e-33;
  Matches  330;  Conservative    0;  Mismatches  210;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcgggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db    915 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 974

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     || ||| || ||    || ||||
Db    975 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCT 1034

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1035 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1094

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1095 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1154

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          |||||     | |  || |||||||||||| || ||| || |||| || || || |   |
Db   1155 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1214

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || |||||  |  |||||||  | |||||  ||  | |
Db   1215 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1274

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| ||  | ||||||||  || || ||||| |||||
Db   1275 AAATAGTTCAAATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTT 1334

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||     |  | ||  | |  || |  ||    |||||||| |||
Db   1335 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1394

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
            |||| | | || ||  |       | |||||  | ||  | || || ||  |  ||
Db   1395 CATTGTCAACCACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAAT 1454

Qy   1698 tac 1700
          |||
Db   1455 TAC 1457


RESULT  25
NMIGAF48
LOCUS       NMIGAF48                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (HF48).
ACCESSION   X82480
VERSION     X82480.1  GI:732875
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="HF48"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57863.1"
                     /db_xref="GI:732876"
                     /db_xref="SPTREMBL:Q57035"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQKADSNQKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHTSTITLTGKSLITNPNSLSVHSI"
BASE COUNT      496 a    319 c    325 g    349 t
```

```
ORIGIN


  Query Match             4.4%;  Score 192.8;  DB 1;  Length 1489;
  Best Local Similarity   60.4%;  Pred. No. 2.8e-33;
  Matches  333;  Conservative    0;  Mismatches  215;  Indels    3;  Gaps    1;

Qy    1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
           | ||| |  | |   |||||||   |  | ||     | ||||||||||||||||||
Db     915 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 974

Qy    1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
           |||| ||| |||  |||| ||||  || |||     |  ||| || ||    || ||||
Db     975 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1034

Qy    1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
           ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db    1035 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1094

Qy    1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  |  |  | | || ||| |  ||
Db    1095 TGGCGACAGATTGGCGAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1154

Qy    1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
           ||||||     | |  || ||||||||||| || ||| || ||||  | || ||    ||
Db    1155 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAA 1214

Qy    1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  ||||||| | |||||  ||  | |
Db    1215 CCAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1274

Qy    1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
            |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db    1275 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1334

Qy    1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db    1335 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1394

Qy    1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |       || |||||  | ||  | || || ||       ||
Db    1395 CATTGTCAACCACAACACAGGCCACACCTCCACAATCACGCTAACGGGTAAATCTTTAAT 1454

Qy    1698 tacacaaccga 1708
           |||  | ||||
Db    1455 TACCAATCCGA 1465


RESULT  26
NMIGAF96
LOCUS       NMIGAF96                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (HF96).
ACCESSION   X82475
VERSION     X82475.1  GI:732879
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 'C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="HF96"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57858.1"
                     /db_xref="GI:732880"
                     /db_xref="SPTREMBL:Q57035"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQKADSNQKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHTSTITLTGKSLITNPNSLSVHSI"
BASE COUNT      496 a    319 c    325 g    349 t
ORIGIN


  Query Match             4.4%;  Score 192.8;  DB 1;  Length 1489;
  Best Local Similarity   60.4%;  Pred. No. 2.8e-33;
  Matches  333;  Conservative    0;  Mismatches  215;  Indels    3;  Gaps    1;

Qy    1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
           | ||| |  | |   |||||||   |  | ||     | ||||||||||||||||||
Db     915 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 974

Qy    1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
           |||| ||| |||  |||| ||||  || |||     |  ||| || ||    || ||||
Db     975 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1034

Qy    1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
           ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db    1035 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1094

Qy    1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  |  |  | | || ||| |  ||
Db    1095 TGGCGACAGATTGGCGAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1154

Qy    1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
           ||||||     | |  || ||||||||||| || ||| || ||||  | || ||    ||
Db    1155 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAA 1214

Qy    1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  ||||||| | |||||  ||  | |
Db    1215 CCAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1274

Qy    1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
            |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db    1275 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1334

Qy    1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db    1335 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1394

Qy    1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |       || |||||  | ||  | || || ||       ||
Db    1395 CATTGTCAACCACAACACAGGCCACACCTCCACAATCACGCTAACGGGTAAATCTTTAAT 1454

Qy    1698 tacacaaccga 1708
           |||  | ||||
Db    1455 TACCAATCCGA 1465


RESULT  27
```

```
NMIGATH2
LOCUS       NMIGATH2                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (ETH2).
ACCESSION   X82469
VERSION     X82469.1  GI:732887
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="ETH2"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57852.1"
                     /db_xref="GI:732888"
                     /db_xref="SPTREMBL:Q57309"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYT
                     VKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI"
BASE COUNT      492 a    321 c    325 g    351 t
ORIGIN
```

```
  Query Match             4.4%;  Score 192.8;  DB 1;  Length 1489;
  Best Local Similarity   60.8%;  Pred. No. 2.8e-33;
  Matches  330;  Conservative    0;  Mismatches  210;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||
Db    915 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 974

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     || ||| || ||    || ||||
Db    975 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCT 1034

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1035 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1094

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1095 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1154

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||| || ||| || |||| || || || |   |
```

```
Db   1155 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1214

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || |||||  |  |||||||  | |||||  ||  | |
Db   1215 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1274

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
             |||      |||||| | || ||||| ||  | ||||||||  || || ||||| |||||
Db   1275 AAATAGTTCAAATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTT 1334

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   1335 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1394

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
             |||| | | || ||  |         | |||||  | ||  | || || ||   |   ||
Db   1395 CATTGTCAACCACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAAT 1454

Qy   1698 tac 1700
             |||
Db   1455 TAC 1457
```

```
RESULT  28
NMIGAG117
LOCUS       NMIGAG117               1525 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NG117).
ACCESSION   X82483
VERSION     X82483.1  GI:732881
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1525)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1525)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1525
                     /organism="Neisseria meningitidis"
                     /strain="NG117"
                     /db_xref="taxon:487"
     gene            1. .1525
                     /gene="iga"
     CDS             <1. .>1525
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57866.1"
                     /db_xref="GI:732882"
                     /db_xref="SPTREMBL:Q51171"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYA
                     VSVKHAKAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNM
                     ARFNKFVTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKG
                     YDLRDLSQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALT
                     NYGVLGDSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDN
                     AGTIKGNREHHWNITFGTNSKIGSTAVRLANNERDANNGQNVTFEDNGTLVLDQNINQ
                     GAGGLFFKGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEIN
                     GTGVNQGQLKVGDGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYF
                     GFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTDRASTITLTGKSLITAPQNLSVYEI
                     "
BASE COUNT      507 a    318 c    334 g    366 t
```

ORIGIN

```
  Query Match             4.4%;  Score 192.8;  DB 1;  Length 1525;
  Best Local Similarity   60.8%;  Pred. No. 2.8e-33;
  Matches  330;  Conservative    0;  Mismatches  210;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db    951 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1010

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     || ||| || ||    || ||||
Db   1011 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCT 1070

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1071 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1130

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  |  |  | | | || ||| |  ||
Db   1131 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1190

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || |||| || || || |   |
Db   1191 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1250

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1251 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1310

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| ||  | |||||||| || || ||||| |||||
Db   1311 AAATAGTTCAAATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTT 1370

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   1371 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1430

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
             |||| | | || ||  |        | |||||  | ||  | || || ||   |   ||
Db   1431 CATTGTCAACCACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAAT 1490

Qy   1698 tac 1700
          |||
Db   1491 TAC 1493
```

RESULT 29
AF012211
LOCUS AF012211 2980 bp DNA linear BCT 30-JAN-2001
DEFINITION Neisseria meningitidis strain B40 IgA1 protease precursor (iga) gene, partial cds.
ACCESSION AF012211
VERSION AF012211.1 GI:2529505
KEYWORDS .
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 2980)
AUTHORS Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del Valle,J. and Achtman,M.
TITLE Clonal descent and microevolution of Neisseria meningitidis during 30 years of epidemic spread
JOURNAL Mol. Microbiol. 25 (6), 1047-1064 (1997)
MEDLINE 98010345
REFERENCE 2 (bases 1 to 2980)
AUTHORS Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del Valle,J. and Achtman,M.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare Genetik, Ihnestrasse 73, Berlin 14195, Germany

```
FEATURES             Location/Qualifiers
     source          1. .2980
                     /organism="Neisseria meningitidis"
                     /strain="B40"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup I; isolated in Morocco, 1967;
                     also present in strain B439, serogroup A, subgroup II,
                     isolated in Djibouti, 1966"
     mat_peptide     1. .>2980
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2980
                     /gene="iga"
                     /allele="iga7"
     CDS             <1. .>2980
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45794.2"
                     /db_xref="GI:12597971"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     GANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGINNDITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQV
                     GIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTG
                     HTSTITLTGKSLITNPNSLSVHSIQNDYDEDDYSYYYRPRRPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAR
                     KDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVSDITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNVNLSQNAALTLGKAALW
                     GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKINH
                     LSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVRDR
                     SHLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQTD
                     SAQIAKPQNIVVAPPSPQAN"
BASE COUNT 985 a 680 c 636 g 679 t
ORIGIN
```

```
  Query Match             4.4%;  Score 192.8;  DB 1;  Length 2980;
  Best Local Similarity   60.4%;  Pred. No. 3.1e-33;
  Matches  333;  Conservative    0;  Mismatches  215;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | |   ||||||||   |  | ||     | ||||||||||||||||||||
Db   1025 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1084

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     |  ||| || ||    || ||||
Db   1085 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1144

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1145 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1204

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  |  |  | | | || ||| |  ||
Db   1205 TGGCGACAGATTGGCGAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1264

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || |||| | || ||    ||
Db   1265 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAA 1324

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1325 CCAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1384

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
```

```
Db   1385 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1444

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   1445 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1504

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |       || |||||  | ||  | || || ||       ||
Db   1505 CATTGTCAACCACAACACAGGCCACACCTCCACAATCACGCTAACGGGTAAATCTTTAAT 1564

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1565 TACCAATCCGA 1575


RESULT  30
AF012207
LOCUS       AF012207                2992 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z3524 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012207
VERSION     AF012207.1  GI:2529497
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2992
                     /organism="Neisseria meningitidis"
                     /strain="Z3524"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup III; isolated in Chad, 1988"
     mat_peptide     1. .>2992
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2992
                     /gene="iga"
                     /allele="iga3"
     CDS             <1. .>2992
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45790.2"
                     /db_xref="GI:12597967"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     GANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGANNDITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQQADADKKVQAFSQV
                     GIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTD
                     RASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGQTENNDWVFMGYKQEEAQKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSGRPTPHARDFVNKSSAR
                     KDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVSNITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFGATQINGNVNLSQNAALVLGKAALW
                     GQIQGQGNSSVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKINH
                     LSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQDR
                     SHLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQKA
                     TQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      986 a    685 c    640 g    681 t
ORIGIN


  Query Match             4.4%;  Score 192.8;  DB 1;  Length 2992;
  Best Local Similarity   60.8%;  Pred. No. 3.1e-33;
  Matches  330;  Conservative    0;  Mismatches  210;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db   1025 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1084

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     || ||| || ||    || ||||
Db   1085 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCT 1144

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1145 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1204

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  |  |  | | || ||| |  ||
Db   1205 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1264

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||      | |  || |||||||||| || ||| || |||| || || || |   |
Db   1265 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1324

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||  | | |
Db   1325 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1384

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
          |||      |||||| | || ||||| ||  | |||||||| || || ||||| |||||
Db   1385 AAATAGTTCAAATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTT 1444

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   1445 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1504

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |       | |||||  | ||  | || || ||   |   ||
Db   1505 CATTGTCAACCACAACACAGACCGCGCCTCCACAATCACATTAACCGGCAAATCATTAAT 1564

Qy   1698 tac 1700
          |||
Db   1565 TAC 1567


RESULT  31
NMIGAF13
LOCUS       NMIGAF13                5261 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (HF13).
ACCESSION   X82474
VERSION     X82474.1  GI:732873
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 5261)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 5261)
  AUTHORS   Lomholt,H.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .5261
                     /organism="Neisseria meningitidis"
                     /strain="HF13"
                     /db_xref="taxon:487"
                     /clone_lib="lambda-phage"
     gene            436. .5121
                     /gene="iga"
     CDS             436. .5121
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57857.1"
                     /db_xref="GI:732874"
                     /db_xref="SPTREMBL:Q51169"
                     /translation="MKTKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
                     KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWSASNLGRLEDYNMARFNKF
                     VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKGYDLRDL
                     SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALTNYGVLG
                     DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDNAGTIKG
                     YGEHHWKTTGTNSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFF
                     KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
                     ELKVGDGTVILNQKADSNQKVSAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
                     DANGNDLTFEHIRNVDEGARIVNHNTSHASTITLTGKSLITNPNSLSVHSIQNDYDED
                     DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGSVNAPMPENGQTENNDWILMGSTQEE
                     AKKNAMNHKNNQRISGFSGFFGEENGKGHNGALNLNFNGKSAQNRFLLTGGTNLNGKI
                     SVTQGNVLLSGRPTPHARDFVNKSSARKDARFSKNNEVVFEDDWINRTFKAAEIAVNQ
                     SASFSSGRNVSNITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTDNLSDKALN
                     SFDATQINGNVNLNQNAALVLGKAALWGQIQGQGNSSVSLNQHSKWHLTSDSQVHNLS
                     LADSHIHLNNASDAQSANKYHTLKINHLSGNGHFHYLTHLAKNLGDKVLVKESASGHY
                     QLHVQDKTGEPNQEGLNLFDASSVQDRSRLSVSLANNHVDLGALRYTIKTENGITRLY
                     NPYAENRRRVKPAPSPATNTASQAQTDSAQIAKPQNIVVAPPSPQANQAEEAKRQQAK
                     AEQVKRQQAEAEKVAHQKAEEAKRQQDALARQQAEQERQRLEAERQAAEIAKQKAEAE
                     EAKRRAAEIAEQKAAAEEAKRQAAELARQQEEARKAAELAAKQKAETERKAAEIAEQK
                     AEAEREAAELAKQKAEEEGRQAAQSQPKRRNRRAIPPELSSDATTRALPRIARNSNPD
                     ASDYEEIPLDALEDEDVSESVDTSDKQPQDNTELHEKVETVSLQPRAAQPRAQAAAQP
                     QAQAVAQADAVSTNTNSALSDAMASTQSILLDTGASLTRHIAQKSRADAEKNSVWMSN
                     IGYGRDYASAQYRRFSSKRTQTQIGIDRSLSENMQIGGVLTYSDSQHTFDQASGKNTF
                     VQANLYGKYYLNDAWYVAGDIGAGSLRSRLQTQQKANFNRTSIQTGLTLGNTLKINQF
                     EIVPSAGIRYSRLSSADYKLGNDSVKVSSMSVKTLTAGLDFAYRFKVGNLTVKPLLSA
                     AYFANYGKGGVNVGGNSFAYKADNQQQYSAGAALLYRNVTLNVNGSITKGKQLEKQKS
                     GQIKIQIRF"
BASE COUNT     1786 a   1235 c   1134 g   1105 t      1 others
ORIGIN
```

```
  Query Match             4.4%;  Score 192.8;  DB 1;  Length 5261;
  Best Local Similarity   60.4%;  Pred. No. 3.3e-33;
  Matches  333;  Conservative    0;  Mismatches  215;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | |   ||||||||   |  | ||     | ||||||||||||||||||||
Db   1539 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1598

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     || ||| || ||    || ||||
Db   1599 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCT 1658

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1659 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1718

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||||  |  |   |  | | || ||| | |||
Db   1719 TGGCGACAGATTGGCGAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTAAA 1778

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          |||||||    | |  || |||||||||||| || ||| || ||||  | || ||    ||
Db   1779 TCAAGGCGAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAA 1838

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||      || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1839 CCAAAAAGTCTCGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1898

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
          |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1899 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1958

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | | || | ||    |||||||| |||
Db   1959 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 2018

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
          |||| | | || ||  | |      | |||||  | ||  | || || ||        ||
Db   2019 CATTGTCAACCACAACACAAGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTAAT 2078

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   2079 TACCAATCCGA 2089
```

```
RESULT  32
AE002424/c
LOCUS       AE002424               10485 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 66 of 206 of
            the complete genome.
ACCESSION   AE002424 AE002098
VERSION     AE002424.1  GI:7225923
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10485)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10485)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10485
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            complement(67. .795)
                     /gene="NMB0696"
     CDS             complement(67. .795)
```

```
                     /gene="NMB0696"
                     /note="similar to GB:M61017 SP:P27675 PID:142988 percent
                     identity: 81.40; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="amino acid ABC transporter, ATP-binding protein"
                     /protein_id="AAF41113.1"
                     /db_xref="GI:7225924"
                     /translation="MIKFKNVHKHFKDLHVINGVNLEIKKGEVVVVCGPSGSGKSTLI
                     RTVNQLESIESGEIWVDGVNVADPKTDLNKIREEVGFVFQGFNLYPHLTVLENITLAP
                     MKVKGQNAEQAEKKAMELLERVGLAHKKDAFPSQLSGGQQQRVAIARGLAMEPRVMLF
                     DEPTSALDPEMVGEVLKVMKDLAESGMTMMCVTHEMGFAREVADRVIFVDKGQILEDE
                     TPEAFFTNPKHERAKQFLQQVMTH"
gene                 complement(968. .1747)
                     /gene="NMB0697"
CDS                  complement(968. .1747)
                     /gene="NMB0697"
                     /note="similar to GB:D10483 SP:P06992 GB:M11054 GB:M68521
                     GB:X04711 percent identity: 69.41; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dimethyladenosine transferase"
                     /protein_id="AAF41114.1"
                     /db_xref="GI:7225925"
                     /translation="MKEHKARKRFGQNFLQDTRIISDIVNAVRPQADDVVIEIGPGLA
                     AITEPLAKKLNRLHVVEIDRDIVCRLKTLPFADKLVIHEGDVLQFDFNGIAGKKKIVG
                     NLPYNISTPLLFKLAEVADDVVDMHFMLQKEVVERMVAAPKSNDYGRLGVMLQYFFDM
                     EMLIDVPPESFDPAPKVDSAVVRMIPVKHRIGKADDFEHFAKLVKLAFHQRRKTIRNN
                     LKELAGDDDLQAVGINPQDRAEHIAPEKYVALSNYLAGKVV"
gene                 1944. .2633
                     /gene="NMB0698"
CDS                  1944. .2633
                     /gene="NMB0698"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41115.1"
                     /db_xref="GI:7225926"
                     /translation="MLFEGLPCLAVWCKNRLSAVLPVGGLNVSENLLEIETHPFDPVL
                     PPKAAVMMMGTFPPKEDKRAMQFHYPNFQNDMWRVYGLVFFNDAAHFQRLSEKAFDAE
                     KIKAFLHERGIASCPTVLKAVRQHGNASDKFLKVVETVDLAAVLAKIPECRHICTTGG
                     KATEILLDIQGGGIKMPKTGETVPFPFAGRDLTLTRLPSTSRAYPLSLAKKAAAYRAF
                     FEMAGLCEKQL"
gene                 2716. .3918
                     /gene="NMB0699"
CDS                  2716. .3918
                     /gene="NMB0699"
                     /note="similar to GB:M58444 SP:P16706 PID:141798
                     PID:461373 percent identity: 80.71; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tryptophan synthase, beta subunit"
                     /protein_id="AAF41116.1"
                     /db_xref="GI:7225927"
                     /translation="MKNYHAPDEKGFFGEHGGLYVSETLIPALQELADAYKAAKNDPE
                     FWEAFRHDLKHYVGRPSPVYHAARLSEHLGGAQIWLKREDLNHTGAHKVNNTIGQALL
                     AKRMGKKRVIAETGAGQHGVASATVAARFGMTCDVYMGADDIQRQMPNVFRMKLLGAN
                     VVGVESGSRTLKDAMNEAMREWVARVDDTFYIIGTAAGPAPYPEMVRDFQCVIGNEAK
                     AQMQEAIGRQPDVAVACVGGGSNAIGLFHPYIGEENVRLVGVEAGGLGVNTPDHAAPI
                     TSGAPIGVLHGFRSYLMQDENGQVLGTHSVSAGLDYPGIGPEHSHLHDIKRVEYTVAK
                     DDEALEAFDLLCRFEGIIPALESSHAVAWAVKNAPKMGKDQVILVNLSGRGDKDINTV
                     AKLKGIKL"
gene                 complement(4097. .9544)
                     /gene="NMB0700"
CDS                  complement(4097. .9544)
                     /gene="NMB0700"
                     /note="similar to GP:732874 percent identity: 93.40;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA-specific serine endopeptidase"
                     /protein_id="AAF41117.1"
                     /db_xref="GI:7225928"
                     /translation="MKTKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHV
                     KGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFV
                     TEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGT
                     PYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDK
                     QEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHHWKTTGT
                     NSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYTVKGINN
                     DITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVIL
                     NQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEH
                     IRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYIQNNDDDDAGYYYYRPRKP
                     IPQGKDLYFKNYRYYALKSGGSVNAPMPENGQTENNDWILMGSTQEEAKKNAMNHKNN
                     QRISGFSGFFGEENGKGHNGALNLNFNGKSAQNRFLLTGGTNLNGKISVTQGNVLLSG
                     RPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKATEIAVNQSASFSSGRNVS
                     DITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALNSFGATQINGNV
                     NLNQNAALVLGKAALWGQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNA
                     SDAQSANKYHTLKINHLSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEP
                     NQEGLNLFDASSVQDRSRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVK
                     PAPSPATNTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEAKRQQAKAEQVKRQ
                     QAEAERKSAELAKQKAEAEREARELATRQKAEQERSSAELARRHEKEREAAELSAKQK
                     VEAEREAQALAVRRKAEAEEAKRQAAELARRHEKEREAAELSAKQRVGEEERRQTAQS
                     QPQRRKRRAAPQDYMAASQDRPKRRGHRSVQQNNVEIAQAQAELARRQQEERKAAELL
                     AKQRAEAEREAQALAARRKAEAEEAKRQAAELAHRQEAERKAAELSANQKAAAEAQAL
                     AARQQKALARQQEEARKAAELAVKQKAETERKTAELAKQRAAAEAAKRQQEARQTAEL
                     ARRQEAERQAAELSAKQKAETDREAAESAKRKAEEEEHRQAAQSQPQRRKRRAAPQDY
                     MAASQNRPKRRGRRSTLPAPPSPSFDSSAYAAPRALHNPDWYENDYEEIPLDALEDEN
                     VSESVDTSDKQPQDNTELHEKYENDYEEIPLDALEDEDVSESVDTSDKQPQDNTELHE
                     KVETVSLQPRAAQPRAQAATQLQAQAAAQADAVSTNTNSALSDAMASTQSILLDTGAS
                     LTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRSLSENMQI
                     GGVLTYSDSQHTFDQASGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSRLQTQQKA
                     NFNRASIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGNDSVKVSSMSVKTLT
                     AGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGNSFVYKADNQQQYSAGAALLY
                     RNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
gene                 complement(9703. .9957)
                     /gene="NMB0701"
CDS                  complement(9703. .9957)
                     /gene="NMB0701"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41118.1"
                     /db_xref="GI:7225929"
                     /translation="MTNPSVRKPAPRHSHEPPSRHYHESGNLGLRVKKTYIPSFPQKQ
                     KTKIRNLKSRHSRKSGNPKTKSRRNLTEEPNPENRLPAFA"
BASE COUNT     2182 a    2461 c    2686 g    3156 t
ORIGIN
```

```
  Query Match              4.4%;  Score 191.4;  DB 1;  Length 10485;
  Best Local Similarity    58.8%;  Pred. No. 7.5e-33;
  Matches  346;  Conservative    0;  Mismatches  239;  Indels    3;  Gaps    1;

Qy    1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
           | ||| |  | |   ||||||||   |  | ||     | ||||||||||||||||||||
Db    8474 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 8415

Qy    1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
           |||| ||| |||  |||| ||||  || |||     |  ||| || ||    || ||||
Db    8414 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 8355

Qy    1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
           ||| |||||  || || |    || | ||  |  ||   |||| |||| ||       |
Db    8354 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 8295

Qy    1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||||  |  |  |  | | || ||| |  ||
Db    8294 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 8235
```

```
Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || ||||||||||| || ||| || |||| || || || |   |
Db   8234 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 8175

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
            |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   8174 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 8115

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   8114 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 8055

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   8054 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 7995

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        | |||||  | ||  | || || ||       ||
Db   7994 CATTGTCAACCACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGAT 7935

Qy   1698 tacacaaccgagtggtaagaatatcaatagacttaattacagcaaaga 1745
           |||  | || |     |    |||  ||    || |  | |   | ||
Db   7934 TACCGATCCCAAAACCATCTCTATTCATTATATTCAAAATAATGATGA 7887
```

RESULT 33
AX044030/c

LOCUS	AX044030	349980 bp	DNA	linear	PAT 24-NOV-2000
DEFINITION	Sequence 109 from Patent WO0066791.				
ACCESSION	AX044030				
VERSION	AX044030.1 GI:11342914				
KEYWORDS	.				
SOURCE	Neisseria meningitidis.				
ORGANISM	Neisseria meningitidis				
	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
REFERENCE	1 (bases 1 to 349980)				
AUTHORS	Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C., Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M., Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.				
TITLE	Neisseria genomic sequences and methods of their use				
JOURNAL	Patent: WO 0066791-A 109 09-NOV-2000; CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)				

FEATURES Location/Qualifiers

source 1. .349980
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
/note="sequence too long, cut in 8 pieces.~seq 1: 1 to 349980 349980 bases~seq 108: 300001 to 649980 349980 bases~seq 109: 600001 to 949980 349980 bases~seq 110: 900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980 349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq 113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to 2272325 172325 bases"

BASE COUNT 82523 a 82940 c 96712 g 87805 t
ORIGIN

Query Match 4.4%; Score 191.4; DB 6; Length 349980;
Best Local Similarity 58.8%; Pred. No. 1.2e-32;
Matches 346; Conservative 0; Mismatches 239; Indels 3; Gaps 1;

```
Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | |   ||||||||   |  | ||     | ||||||||||||||||||
Db 130048 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 129989

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     |  ||| || ||    || ||||
Db 129988 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 129929

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db 129928 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 129869

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
            |||| |  || | ||||||||||||||||||  |  |  |  | | || ||| |  ||
Db 129868 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 129809

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || ||||||||||| || ||| || |||| || || || |   |
Db 129808 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 129749

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
            |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db 129748 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 129689

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||| || || |||||||| || || ||||| |||||
Db 129688 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 129629

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db 129628 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 129569

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        | |||||  | ||  | || || ||       ||
Db 129568 CATTGTCAACCACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGAT 129509

Qy   1698 tacacaaccgagtggtaagaatatcaatagacttaattacagcaaaga 1745
           |||  | || |     |    |||  ||    || |  | |   | ||
Db 129508 TACCGATCCCAAAACCATCTCTATTCATTATATTCAAAATAATGATGA 129461
```

RESULT 34
NMIGA139

LOCUS	NMIGA139	1489 bp	DNA	linear	BCT 14-MAR-1995
DEFINITION	N.meningitidis iga gene (81139).				
ACCESSION	X82477				
VERSION	X82477.1 GI:732856				
KEYWORDS	iga gene; IgA1 protease.				
SOURCE	Neisseria meningitidis.				
ORGANISM	Neisseria meningitidis				
	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
REFERENCE	1 (bases 1 to 1489)				
AUTHORS	Lomholt,H., Poulsen,K. and Kilian,M.				
TITLE	Comparative characterization of the iga gene encoding IgA1 protease in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus influenzae				
JOURNAL	Mol. Microbiol. 15 (3), 495-506 (1995)				
MEDLINE	95302961				
REFERENCE	2 (bases 1 to 1489)				
AUTHORS	Lomholt,H.				
TITLE	Direct Submission				
JOURNAL	Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology, The Bartholin Building, University of Aarhus, 8000 C Aarhus C, DENMARK				

FEATURES Location/Qualifiers

source 1. .1489
/organism="Neisseria meningitidis"
/strain="81139"
/db_xref="taxon:487"

gene 1. .1489
/gene="iga"

CDS <1. .>1489
/gene="iga"
/codon_start=1
/transl_table=11
/product="IgA1 protease"
/protein_id="CAA57860.1"
/db_xref="GI:732857"
/db_xref="SPTREMBL:Q51163"
/translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR

```
                    YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                    LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHH
                    WKTTGTNSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                    VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                    DGTVILNQQADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGN
                    DLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYI"
BASE COUNT      498 a    320 c    321 g    350 t
ORIGIN


  Query Match             4.4%;  Score 191.2;  DB 1;  Length 1489;
  Best Local Similarity   60.3%;  Pred. No. 6.5e-33;
  Matches  332;  Conservative    0;  Mismatches  216;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | |  |||||||   |  | ||     | ||||||||||||||||||||
Db    915 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 974

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     |  ||| || ||    || ||||
Db    975 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1034

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||       |
Db   1035 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1094

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1095 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1154

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||| || ||| || |||| || || || |   |
Db   1155 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1214

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||. | |
Db   1215 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1274

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1275 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1334

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    | | ||  | |  || |  ||    |||||||| |||
Db   1335 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1394

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |      | |||||  | ||  | || || ||       ||
Db   1395 CATTGTCAACCACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGAT 1454

Qy   1698 tacacaaccga 1708
          |||  | || |
Db   1455 TACCGATCCCA 1465


RESULT  35
NMIGA894
LOCUS       NMIGA894                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (SM1894).
ACCESSION   X82476
VERSION     X82476.1  GI:732867
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961

REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="SM1894"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57859.1"
                     /db_xref="GI:732868"
                     /db_xref="SPTREMBL:Q51166"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHH
                     WKTTGTNSHIGSTAVRLAGNERGANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQKADSNQKVQAFSQVGIVSGRRTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHTSTITLTGKSLITNPNSLSVHSI"
BASE COUNT      496 a    320 c    324 g    349 t
ORIGIN


  Query Match             4.4%;  Score 191.2;  DB 1;  Length 1489;
  Best Local Similarity   60.3%;  Pred. No. 6.5e-33;
  Matches  332;  Conservative    0;  Mismatches  216;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | |  |||||||   |  | ||     | ||||||||||||||||||||
Db    915 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 974

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     |  ||| || ||    || ||||
Db    975 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1034

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||       |
Db   1035 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1094

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1095 TGGCGACAGATTGGCGAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1154

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||| || ||| || |||| | || ||    ||
Db   1155 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAA 1214

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | ||||  ||  | |
Db   1215 CCAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCCGTACATTGGTATT 1274

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1275 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1334

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    | | ||  | |  || |  ||    |||||||| |||
Db   1335 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1394

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        || |||||  | ||  | || || ||       ||
Db   1395 CATTGTCAACCACAACACAGGCCACACCTCCACAATCACGCTAACGGGTAAATCTTTAAT 1454
```

```
Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1455 TACCAATCCGA 1465


RESULT  36
NMIGAC16
LOCUS       NMIGAC16                1525 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NGC16).
ACCESSION   X82485
VERSION     X82485.1  GI:732869
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1525)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1525)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1525
                     /organism="Neisseria meningitidis"
                     /strain="NGC16"
                     /db_xref="taxon:487"
     gene            1. .1525
                     /gene="iga"
     CDS             <1. .>1525
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57868.1"
                     /db_xref="GI:732870"
                     /db_xref="SPTREMBL:Q51167"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYA
                     VSVKHAKAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNM
                     ARFNKFVTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKG
                     YDLRDLSQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALT
                     NYGVLGDSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDN
                     AGTIKGNGEHHWNITFGTNSHIGSTAVRLAGNERDANNGQNVTFEDNGTLVLDQNINQ
                     GAGGLFFKGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEIN
                     GTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYF
                     GFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYI
                     "
BASE COUNT      504 a    322 c    336 g    363 t
ORIGIN


  Query Match             4.4%;  Score 191.2;  DB 1;  Length 1525;
  Best Local Similarity   60.3%;  Pred. No. 6.5e-33;
  Matches  332;  Conservative    0;  Mismatches  216;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||
Db    951 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1010

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||     || ||| || ||    || ||||
Db   1011 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCT 1070

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |

Db   1071 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1130

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1131 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1190

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || ||||  | || ||    ||
Db   1191 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAA 1250

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  ||||||  | |||||  ||  | |
Db   1251 CCAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1310

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
            |||     |||||| | || ||||| ||  | |||||||| || || ||||| |||||
Db   1311 AAATAGTTCAAATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTT 1370

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | |  || |  ||    ||||||| |||
Db   1371 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1430

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
            |||| | | || ||  |        | |||||  | ||  | || || ||       ||
Db   1431 CATTGTCAACCACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGAT 1490

Qy   1698 tacacaaccga 1708
          |||  | || |
Db   1491 TACCGATCCCA 1501


RESULT  37
AF012206
LOCUS       AF012206                2992 bp    DNA     linear   BCT 30-JAN-2001
DEFINITION  Neisseria meningitidis strain Z4024 IgA1 protease precursor (iga)
            gene, partial cds.
ACCESSION   AF012206
VERSION     AF012206.1  GI:2529495
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Muller,K., Seiler,A., Wang,J.F., del
            Valle,J. and Achtman,M.
  TITLE     Clonal descent and microevolution of Neisseria meningitidis during
            30 years of epidemic spread
  JOURNAL   Mol. Microbiol. 25 (6), 1047-1064 (1997)
  MEDLINE   98010345
REFERENCE   2  (bases 1 to 2992)
  AUTHORS   Morelli,G., Malorny,B., Mueller,K., Seiler,A., Wang,J.-F., del
            Valle,J. and Achtman,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Max-Planck Institut fuer Molekulare
            Genetik, Ihnestrasse 73, Berlin 14195, Germany
FEATURES             Location/Qualifiers
     source          1. .2992
                     /organism="Neisseria meningitidis"
                     /strain="Z4024"
                     /db_xref="taxon:487"
                     /note="serogroup A, subgroup VI; isolated in East-Germany,
                     1985"
     mat_peptide     1. .>2992
                     /gene="iga"
                     /product="IgA1 protease"
     gene            <1. .>2992
                     /gene="iga"
                     /allele="iga6"
     CDS             <1. .>2992
                     /gene="iga"
                     /codon_start=3
                     /transl_table=11
```

```
                     /product="IgA1 protease precursor"
                     /protein_id="AAC45789.2"
                     /db_xref="GI:12597966"
                     /translation="LAYALTPYSEAALVRDDVDYQIFRDFAENKGKFFVGATDLSVKN
                     KQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYAVSVKHVKGDEISYYGHHNGHLD
                     VSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMARLNKFVTEVAPIAPTSAGGGVE
                     TYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYRYAIAGTPYQDVNVTSNLNQEGL
                     IGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSPLFAYDKQEKRWVFLGAYDYWAG
                     YQKNSWQEWNIYKKEFADEIKQRDNAGTIKGYGEHHWKTTGTNSHIGSTAVRLAGNER
                     DANNGQNVTFEDNGTLVLDQNINQGAGGLFFKGDYTVKGANNGITWLGAGIDVADGKK
                     VVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQV
                     GIVSGRGTLVLNSSNQINPDNLYFGFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTS
                     HASTITLTGKSLITAPQNLSVYEIRNDYDDDDYYGYYSYRKPIPQGKDLYYKNYRYYA
                     LKSGGSVNAPMPENGVTENNDWVFMGYTQEEAKKNAMNHKNNQRISGFSGFFGEENGK
                     GHNGALNLNFNGKSAQNRFLLTGGANLNGKISVTQGNVLLSGRPTPHARDFVNKSSAR
                     KDAHFSKNNEVVFEDDWINRTFKAAEITVNQSASFSSGRNVSDITANITATDNAKVNL
                     GYKNGDEVCVRSDYTGYVTCNTDNLSDKALNSFGATQINGNVNLSQNAALVLGKAALW
                     GQIQGQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYHTLKINH
                     LSGNGHFHYLTHLAKNLGDKVLVKESASGHYQLHVQDKTGEPNQEGLNLFDASSVQDR
                     SRLSVSLANNHVDLGALRYTIKTENGITRLYNPYAENRRRVKPAPSPATNTASQAQKA
                     TQTDGAQIAKPQNIVVAPPSPQAN"
BASE COUNT      983 a    684 c    639 g    686 t
ORIGIN


  Query Match             4.4%;  Score 191.2;  DB 1;  Length 2992;
  Best Local Similarity   60.6%;  Pred. No. 7.1e-33;
  Matches  329;  Conservative    0;  Mismatches  211;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db   1025 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1084

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaa---cgaaacgtggca 1277
          |||| ||| ||| |||| ||||  || |||     || ||| ||   |   || ||||
Db   1085 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGGCATCACTTGGCT 1144

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1145 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1204

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1205 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1264

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || ||||||||||| || ||| || ||||  | || ||    ||
Db   1265 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATTCAAA 1324

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  ||||||  | |||||  ||  | |
Db   1325 CCAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1384

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
             |||      |||||| | || ||||| ||  | ||||||| || || ||||| ||||
Db   1385 AAATAGTTCAAATCAGATTAATCCCGATAACTTATATTTCGGTTTCCGTGGCGGTCGTTT 1444

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | |  || |  ||    ||||||| |||
Db   1445 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1504

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  | |      | |||||  | ||  | || || ||   |   ||
Db   1505 CATTGTCAACCACAACACAAGCCACGCCTCCACAATCACATTAACCGGCAAATCATTAAT 1564

Qy   1698 tac 1700
          |||
Db   1565 TAC 1567


RESULT  38
NGIGAPRO
LOCUS       NGIGAPRO                4899 bp    DNA     linear   BCT 21-MAR-1995
DEFINITION  Neisseria gonorrhoeae gene for IgA protease.
ACCESSION   X04835
VERSION     X04835.1  GI:44868
KEYWORDS    IgA protease; inverted repeat; protease; secreted protein; signal
            peptide.
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   Pohlner,J., Halter,R., Beyreuther,K. and Meyer,T.F.
  TITLE     Gene structure and extracellular secretion of Neisseria gonorrhoeae
            IgA protease
  JOURNAL   Nature 325 (6103), 458-462 (1987)
  MEDLINE   87115823
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="Neisseria gonorrhoeae"
                     /strain="M11"
                     /db_xref="taxon:485"
                     /clone="pIP100"
     promoter        34. .39
                     /note="put. -43 region"
     promoter        43. .48
                     /note="put. -35 region"
     promoter        63. .68
                     /note="put. -10 region"
     RBS             87. .97
                     /note="put. rRNA binding region"
     CDS             104. .4702
                     /note="IgA protease precursor; 169K (AA -27 to 1505)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA28538.1"
                     /db_xref="GI:44869"
                     /db_xref="SWISS-PROT:P09790"
                     /translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
                     KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
                     VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
                     SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
                     DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
                     NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
                     KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
                     QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
                     DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
                     DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
                     ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQKRFLLTGGANLNGKI
                     SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
                     SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
                     SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
                     LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
                     QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
                     NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
                     QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
                     ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
                     IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
                     STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
                     LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
                     LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
                     LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
                     MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
                     AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
     sig_peptide     104. .184
     mat_peptide     185. .4699
                     /product="mature IgA protease (AA 1-1505)"
     misc_feature    3061. .3062
                     /note="autoproteolytic site for IgA protease"
     misc_feature    3157. .3158
                     /note="autoproteolytic site for IgA protease"
     misc_feature    3466. .3467
                     /note="autoproteolytic site for IgA protease"
     repeat_unit     4725. .4739
```

```
                     /note="inverted repeat A"
     terminator      4725. .4762
                     /note="put. transcription termination signal"
     repeat_unit     4747. .4762
                     /note="inverted repeat A'"
BASE COUNT     1617 a   1104 c   1080 g   1098 t
ORIGIN
```

Query Match 4.4%; Score 191.2; DB 1; Length 4899;
Best Local Similarity 60.3%; Pred. No. 7.6e-33;
Matches 332; Conservative 0; Mismatches 216; Indels 3; Gaps 1;

```
Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db   1207 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGAATCAAAACATCAACCAAGGCGCGGG 1266

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||  ||| |||  |||| ||||  || |||      | ||| || ||    || |||
Db   1267 CGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGCAAATAATGACATCACTTGGTT 1326

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| ||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1327 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1386

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||||  || |  |  | | || ||| |  ||
Db   1387 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGAAATAAACGGCACAGGCGTTAA 1446

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || ||||  | || || |   |
Db   1447 CCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATGCCGA 1506

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1507 CAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1566

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || |||||  |||||
Db   1567 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1626

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   1627 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1686

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        | |||||  | ||  | || || ||       ||
Db   1687 CATTGTCAACCACAACACAGACCACGCCTCCACAATCACACTAACGGGTAAATCTTTAAT 1746

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1747 TACCAATCCGA 1757
```

```
RESULT  39
A02796
LOCUS       A02796                  4899 bp    DNA     linear   PAT 10-MAY-1993
DEFINITION  N.gonorrhoeae gene for IgA-protease precursor.
ACCESSION   A02796
VERSION     A02796.1  GI:345152
KEYWORDS    IgA protease.
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   Meyer,T.F., Halter,R. and Pohlner,J.
  TITLE     Process for recovering proteins by means of gene technology using
            gram-negative host cells
  JOURNAL   Patent: EP 0254090-A 1 27-JAN-1988;
            Max-Planck-Gesellschaft zur Foerderung der Wissenschaften
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
     CDS             104. .4702
                     /codon_start=1
                     /transl_table=11
                     /product="IgA protease"
                     /protein_id="CAA00270.1"
                     /db_xref="GI:345153"
                     /translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
                     KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
                     VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
                     SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
                     DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
                     NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
                     KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
                     QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
                     DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
                     DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
                     ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQNRFLLTGGANLNGKI
                     SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
                     SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
                     SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
                     LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
                     QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
                     NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
                     QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
                     ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
                     IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
                     STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
                     LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
                     LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
                     LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
                     MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
                     AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
BASE COUNT     1617 a   1105 c   1079 g   1098 t
ORIGIN
```

Query Match 4.4%; Score 191.2; DB 6; Length 4899;
Best Local Similarity 60.3%; Pred. No. 7.6e-33;
Matches 332; Conservative 0; Mismatches 216; Indels 3; Gaps 1;

```
Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db   1207 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGAATCAAAACATCAACCAAGGCGCGGG 1266

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||  ||| |||  |||| ||||  || |||      | ||| || ||    || |||
Db   1267 CGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGCAAATAATGACATCACTTGGTT 1326

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| ||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1327 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1386

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||||  || |  |  | | || ||| |  ||
Db   1387 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGAAATAAACGGCACAGGCGTTAA 1446

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || ||||  | || || |   |
Db   1447 CCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATGCCGA 1506

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1507 CAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1566

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || |||||  |||||
Db   1567 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1626

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
```

```
          ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   1627 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1686

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        | |||||  | ||  | || || ||       ||
Db   1687 CATTGTCAACCACAACACAGACCACGCCTCCACAATCACACTAACGGGTAAATCTTTAAT 1746

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1747 TACCAATCCGA 1757


RESULT  40
A12416
LOCUS       A12416                  4899 bp    DNA     linear   PAT 05-OCT-1994
DEFINITION  IgA-protease gene precursor.
ACCESSION   A12416
VERSION     A12416.1  GI:640585
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   .
  TITLE     PROCESS FOR THE GENETIC PRODUCTION OF PROTEINS BY USING OF
            GRAM-NEGATIVE HOST CELLS
  JOURNAL   Patent: WO 8800238-A 12 14-JAN-1988;
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="Neisseria gonorrhoeae"
                     /db_xref="taxon:485"
     gene            104. .4702
                     /gene="IgA protease precursor"
     CDS             104. .4702
                     /gene="IgA protease precursor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA01030.1"
                     /db_xref="GI:640586"
                     /translation="MKAKRFKINAISLSIFLAYALTPYSEAALVRDDVDYQIFRDFAE
                     NKGKFFVGATDLSVKNKRGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYAVSVKHA
                     KAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWGASNLGRLEDYNMARFNKF
                     VTEVAPIAPTDAGGGLDTYKDKNRFSSFVRIGAGRQLVYEKGVYHQEGNEKGYDLRDL
                     SQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNKQYSAEELKQALSQDALTNYGVLG
                     DSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQHDNAGTVKG
                     NGEHHWKTTGTNSHIGSTAVRLANNEGDANNGQNVTFEDNGTLVLNQNINQGAGGLFF
                     KGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQG
                     QLKVGDGTVILNQKADADKKVQAFSQVGIVSGRGTLVLNSSNQINPDNLYFGFRGGRL
                     DANGNDLTFEHIRNVDEGARIVNHNTDHASTITLTGKSLITNPNSLSVHSIQNDYDED
                     DYSYYYRPRRPIPQGKDLYYKNYRYYALKSGGRLNAPMPENGVAENNDWIFMGYTQEE
                     ARKNAMNHKNNRRIGDFGGFFDEENGKGHNGALNLNFNGKSAQNRFLLTGGANLNGKI
                     SVTQGNVLLSGRPTPHARDFVNKSSARKDAHFSKNNEVVFEDDWINRTFKAAEIAVNQ
                     SASFSSGRNVSDITANITATDNAKVNLGYKNGDEVCVRSDYTGYVTCNTGNLSDKALN
                     SFDATRINGNVNLNQNAALVLGKAALWGKIQGQGNSRVSLNQHSKWHLTGDSQVHNLS
                     LADSHIHLNNASDAQSANKYHTIKINHLSGNGHFHYLTDLAKNLGDKVLVKESASGHY
                     QLHVQNKTGEPNQEGLDLFDASSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLY
                     NPYAGNGRPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVVAPPSPQANQAEEALR
                     QQAKAEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEARKAAELAAKQKAEAERK
                     ARELARQKAEEASHQANAKPKRRRRRAILPRPPAPVFSLDDYDAKDNSESSIGNLARV
                     IPRMGRELINDYEEIPLEELEDEAEEERRQATQFHSKSRNRRAISSEPSSDEDASESV
                     STSDKHPQDNTELHEKVETAGLQPRAAQPRTQAAAQADAVSTNTNSALSDAMASTQSI
                     LLDTGAYLTRHIAQKSRADAEKNSVWMSNTGYGRDYASAQYRRFSSKRTQTQIGIDRS
                     LSENMQIGGVLTYSDSQHTFDQAGGKNTFVQANLYGKYYLNDAWYVAGDIGAGSLRSR
                     LQTQQKANFNRTSIQTGLTLGNTLKINQFEIVPSAGIRYSRLSSADYKLGDDSVKVSS
                     MAVKTLTAGLDFAYRFKVGNLTVKPLLSAAYFANYGKGGVNVGGKSFAYKADNQQQYS
                     AGVALLYRNVTLNVNGSITKGKQLEKQKSGQIKIQIRF"
BASE COUNT     1617 a   1105 c   1079 g   1098 t
ORIGIN


  Query Match             4.4%;  Score 191.2;  DB 6;  Length 4899;
  Best Local Similarity   60.3%;  Pred. No. 7.6e-33;
  Matches  332;  Conservative    0;  Mismatches  216;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | |||||||||   |  | ||     | ||||||||||||||||||||
Db   1207 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGAATCAAAACATCAACCAAGGCGCGGG 1266

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
           |||  ||| |||  |||| ||||  || |||      | ||| || ||    || |||
Db   1267 CGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGCAAATAATGACATCACTTGGTT 1326

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
           ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1327 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1386

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
             |||| |  || | ||||||||||||||||  || |  |  | | || ||| |  ||
Db   1387 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGAAATAAACGGCACAGGCGTTAA 1446

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
           ||||||     | |  || |||||||||| || ||| || |||  | || || |   |
Db   1447 CCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATGCCGA 1506

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1507 CAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1566

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || |||| || || |||||||| || ||  ||||| |||||
Db   1567 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1626

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   1627 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1686

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        | |||||  | ||  | || || ||       ||
Db   1687 CATTGTCAACCACAACACAGACCACGCCTCCACAATCACACTAACGGGTAAATCTTTAAT 1746

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1747 TACCAATCCGA 1757


RESULT  41
AR079202
LOCUS       AR079202                4899 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 5 from patent US 5965424.
ACCESSION   AR079202
VERSION     AR079202.1  GI:10005948
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4899)
  AUTHORS   Ambrosius,D., Dony,C. and Rudolph,R.
  TITLE     Methods for making neisseria or hemophilus IgA protease and DNA
            encoding the proteases
  JOURNAL   Patent: US 5965424-A 5 12-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .4899
                     /organism="unknown"
BASE COUNT     1617 a   1104 c   1080 g   1098 t
ORIGIN


  Query Match             4.4%;  Score 191.2;  DB 6;  Length 4899;
  Best Local Similarity   60.3%;  Pred. No. 7.6e-33;
  Matches  332;  Conservative    0;  Mismatches  216;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | |||||||||   |  | ||     | ||||||||||||||||||||
Db   1207 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGAATCAAAACATCAACCAAGGCGCGGG 1266
```

```
Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          ||| ||| ||| |||| |||| || |||       | ||| || ||    || |||
Db   1267 CGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGCAAATAATGACATCACTTGGTT 1326

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1327 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1386

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
            |||| |  || | ||||||||||||||||  || |  |  | | || ||| |. ||
Db   1387 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGAAATAAACGGCACAGGCGTTAA 1446

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||| || ||| || ||||  | || || |   |
Db   1447 CCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATGCCGA 1506

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  ||||||| | |||||  ||  | |
Db   1507 CAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1566

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1567 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1626

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||     |  | ||  | |  || |  ||    |||||||| |||
Db   1627 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1686

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        | |||||  | ||  | || || ||      ||
Db   1687 CATTGTCAACCACAACACAGACCACGCCTCCACAATCACACTAACGGGTAAATCTTTAAT 1746

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   1747 TACCAATCCGA 1757
```

```
RESULT  42
NGCOMLGN
LOCUS       NGCOMLGN               8393 bp    DNA     linear   BCT 18-NOV-1996
DEFINITION  N.gonorrhoeae gene for competence lipoprotein ComL.
ACCESSION   Z49895
VERSION     Z49895.1  GI:1107832
KEYWORDS    ComL; competence lipoprotein.
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 8393)
  AUTHORS   Fussenegger,M., Facius,D., Meier,J. and Meyer,T.F.
  TITLE     A novel peptidoglycan-linked lipoprotein (ComL) that functions in
            natural transformation competence of Neisseria gonorrhoeae
  JOURNAL   Mol. Microbiol. 19 (5), 1095-1105 (1996)
  MEDLINE   96249702
REFERENCE   2  (bases 1 to 8393)
  AUTHORS   Fussenegger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1995) Martin Fussenegger Dr., Infektionsbiologie,
            Max-Planck-Institut, fuer Biologie, Spemannstr. 34, Tuebingen, BW,
            D-72076, Germany
FEATURES             Location/Qualifiers
     source          1. .8393
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11A"
                     /db_xref="taxon:485"
     CDS             120. .923
                     /codon_start=1
                     /transl_table=11
                     /product="ComL, competence lipoprotein"
                     /protein_id="CAA90076.1"
                     /db_xref="GI:1107833"
                     /db_xref="SPTREMBL:Q50985"
                     /translation="MKKILLTVSLGLALSACATQGTADKDAQITQDWSVEKLYAEAQD
                     ELNSSNYTRAVKLYEILESRFPTSRHARQSQLDTAYAYYKDDEKDKALAAIERFRRLH
                     PQHPNMDYALYLRGLVLFNEDQSFLNKLASQDWSDRDPKANREAYQAFAELVQRFPNS
                     KYAADATARMVKLVDALGGNEMSVARYYMKRGAYIAAANRAKKIIGSYQNTRYVEESL
                     AILELAYKKLDKPQLAADTRRVLETNFPKSPFLTHAWQPDDMPWWRYWH"
     terminator      934. .967
BASE COUNT     2550 a   2300 c   1886 g   1657 t
ORIGIN
```

```
  Query Match             4.4%;  Score 191.2;  DB 1;  Length 8393;
  Best Local Similarity   60.3%;  Pred. No. 8.1e-33;
  Matches  332;  Conservative    0;  Mismatches  216;  Indels    3;  Gaps    1;

Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
          | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db   4701 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGAATCAAAACATCAACCAAGGCGCGGG 4760

Qy   1221 cggtttgtatttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||  ||| |||  |||| ||||  || |||       | ||| || ||    || |||
Db   4761 CGGCCTGTTTTTCAAAGGCGATTACACAGTCAAAGGCGCAAATAATGACATCACTTGGTT 4820

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   4821 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 4880

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
            |||| |  || | ||||||||||||||||  || |  |  | | || ||| |. ||
Db   4881 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACTTTGGAAATAAACGGCACAGGCGTTAA 4940

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||| || ||| || ||||  | || || |   |
Db   4941 CCAAGGGCAATTAAAGGTCGGCGACGGTACGGTTATTCTGAATCAAAAAGCCGATGCCGA 5000

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||   || || ||     || ||||| |  ||||||| | |||||  ||  | |
Db   5001 CAAAAAAGTTCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 5060

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   5061 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 5120

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||     |  | ||  | |  || |  ||    |||||||| |||
Db   5121 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 5180

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        | |||||  | ||  | || || ||      ||
Db   5181 CATTGTCAACCACAACACAGACCACGCCTCCACAATCACACTAACGGGTAAATCTTTAAT 5240

Qy   1698 tacacaaccga 1708
          |||  | ||||
Db   5241 TACCAATCCGA 5251
```

```
RESULT  43
NMIGA341
LOCUS       NMIGA341               1489 bp    DNA     linear   BCT 12-JUL-1999
DEFINITION  N.meningitidis iga gene (IHN5341).
ACCESSION   X82478
VERSION     X82478.1  GI:732862
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
```

```
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="IHN5341"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /pseudo
                     /transl_table=11
                     /codon_start=1
                     /product="IgA1 protease"
                     /db_xref="PID:e123968"
BASE COUNT      498 a    319 c    321 g    351 t
ORIGIN
```

Query Match 4.4%; Score 189.6; DB 1; Length 1489;
Best Local Similarity 60.1%; Pred. No. 1.5e-32;
Matches 331; Conservative 0; Mismatches 217; Indels 3; Gaps 1;

```
Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcgggg 1220
          | ||| |  | |  ||||||||   |  | ||     | ||||||||||||||||||||
Db    915 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 974

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||      |  ||| || ||    || ||||
Db    975 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1034

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1035 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1094

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
            |||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1095 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1154

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || |||| || || || |   |
Db   1155 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1214

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||    | || ||     || ||||| |  |||||||  | |||||  ||  | |
Db   1215 CAAAAAAGTCTAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCGGTACATTGGTATT 1274

Qy   1518 gaatgccgataatcagttcaaccccgacaaactctatttcggctttcgcggcggacgttt 1577
             |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db   1275 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1334

Qy   1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
          ||||   || ||  ||    |  | ||  | |  || |  ||    |||||||| |||
Db   1335 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1394

Qy   1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           |||| | | || ||  |        | |||||  | ||  | || || ||       ||
Db   1395 CATTGTCAACCACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGAT 1454

Qy   1698 tacacaaccga 1708
          |||  | || |
Db   1455 TACCGATCCCA 1465
```

```
RESULT  44
NMIGAG80
LOCUS       NMIGAG80                1489 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NCG80).
ACCESSION   X82479
VERSION     X82479.1  GI:732883
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1489)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1489
                     /organism="Neisseria meningitidis"
                     /strain="NGC80"
                     /db_xref="taxon:487"
     gene            1. .1489
                     /gene="iga"
     CDS             <1. .>1489
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57862.1"
                     /db_xref="GI:732884"
                     /db_xref="SPTREMBL:Q51172"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNRRTLTVIDPQYA
                     VSVKHVKGDEISYYGHHNGHLDVSNDENEYRSVAQNDYEPNKNWHHGNQGRLEDYNMA
                     RLNKFVTEVAPIAPTSAGGGVETYKDKNRFSEFVRVGAGTQFEYNSRYNMTELSRAYR
                     YAIAGTPYQDVNVTSNLNQEGLIGFGDNSKHHSPEKLKEVLSQNALTNYAVLGDSGSP
                     LFAYDKQEKRWVFLGAYDYWAGYQKNSWQEWNIYKKEFADEIKQRDNAGTIKGNGEHH
                     WKTTGTNSHIGSTAVRLANNERDANNGQNVTFENNGTLVLDQNINQGAGGLFFKGDYT
                     VKGINNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEINGTGVNQGQLKVG
                     DGTVILNQQADADKKVQAFSQVGIVSGRRTLVLNSSNQINPDNLYFGFRGGRLDANGN
                     DLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITDPKTISIHYI"
BASE COUNT      498 a    321 c    320 g    350 t
ORIGIN
```

Query Match 4.4%; Score 189.6; DB 1; Length 1489;
Best Local Similarity 60.1%; Pred. No. 1.5e-32;
Matches 331; Conservative 0; Mismatches 217; Indels 3; Gaps 1;

```
Qy   1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcgggg 1220
          | ||| |  | |  ||||||||   |  | ||     | ||||||||||||||||||||
Db    915 TGTTACCTTTGAAAACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 974

Qy   1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
          |||| ||| |||  |||| ||||  || |||      |  ||| || ||    || ||||
Db    975 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTATAAATAATGACATCACTTGGCT 1034

Qy   1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
          ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db   1035 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1094

Qy   1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
            |||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db   1095 TGGCGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1154

Qy   1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
          ||||||     | |  || |||||||||||| || ||| || |||| || || || |   |
Db   1155 CCAAGGGCAATTAAAAGTCGGCGACGGTACGGTTATTCTGAATCAACAAGCCGATGCCGA 1214

Qy   1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
             |||    | || ||     || ||||| |  |||||||  | |||||  ||  | |
```

```
Db    1215 CAAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCCGTACATTGGTATT 1274

Qy    1518 gaatgccgataatcagttcaacccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db    1275 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1334

Qy    1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||    |  | ||  | |  || |  ||    ||||||||  |||
Db    1335 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1394

Qy    1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           ||||  |  | || ||  |         | |||||  | ||  | || || ||       ||
Db    1395 CATTGTCAACCACAACACAGGCCACGCCTCCACAATCACGCTAACGGGTAAATCTTTGAT 1454

Qy    1698 tacacaaccga 1708
           |||  |  || |
Db    1455 TACCGATCCCA 1465


RESULT  45
NMIGAC65
LOCUS       NMIGAC65                1525 bp    DNA     linear   BCT 14-MAR-1995
DEFINITION  N.meningitidis iga gene (NGC65).
ACCESSION   X82484
VERSION     X82484.1  GI:732871
KEYWORDS    iga gene; IgA1 protease.
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1525)
  AUTHORS   Lomholt,H., Poulsen,K. and Kilian,M.
  TITLE     Comparative characterization of the iga gene encoding IgA1 protease
            in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
            influenzae
  JOURNAL   Mol. Microbiol. 15 (3), 495-506 (1995)
  MEDLINE   95302961
REFERENCE   2  (bases 1 to 1525)
  AUTHORS   Lomholt,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1994) H. Lomholt, Dept.of Medical Microbiology,
            The Bartholin Building, University of Aarhus, 8000 C Aarhus C,
            DENMARK
FEATURES             Location/Qualifiers
     source          1. .1525
                     /organism="Neisseria meningitidis"
                     /strain="NGC65"
                     /db_xref="taxon:487"
     gene            1. .1525
                     /gene="iga"
     CDS             <1. .>1525
                     /gene="iga"
                     /codon_start=1
                     /transl_table=11
                     /product="IgA1 protease"
                     /protein_id="CAA57867.1"
                     /db_xref="GI:732872"
                     /db_xref="SPTREMBL:Q51168"
                     /translation="ATDLSVKNKQGQNIGNALSNVPMIDFSVADVNKRIATVVDPQYA
                     VSVKHAKAEVHTFYYGQYNGHNDVADKENEYRVVEQNNYEPHKAWSASNLGRLEDYNM
                     ARFNKFVTEVAPIAPTDAGGGLDTYKDKNRFSSFVRVGAGRQLVYEKGAYHQEGNEKG
                     YDLRDLSQAYRYAIAGTPYKDINIDQTMNTEGLIGFGNHNTHYSAEELKQALSQDALT
                     NYGVLGDSGSPLFAFDKQKNQWVFLGTYDYWAGYGKKSWQEWNIYKKEFADKIKQRDN
                     AGTIKGNGEHHWNITFGTNSHIGSTAVRLAGNEKDANNGQNVTFEDNGTLVLDQNINQ
                     GAGGLFFKGDYTVKGANNDITWLGAGIDVADGKKVVWQVKNPNGDRLAKIGKGTLEIN
                     GTGVNQGQLKVGDGTVILNQKADSNQKVQAFSQVGIVSGRRTLVLNSSNQINPDNLYF
                     GFRGGRLDANGNDLTFEHIRNVDEGARIVNHNTGHASTITLTGKSLITAPQNLSVYEI
                     "
BASE COUNT      510 a    320 c    330 g    365 t
ORIGIN


  Query Match             4.4%;  Score 189.6;  DB 1;  Length 1525;
  Best Local Similarity   60.4%;  Pred. No. 1.5e-32;
  Matches  328;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy    1161 ttttatcgattacggcaacggcaaactcatcttatcaaacaacatcaaccaaggcgcggg 1220
           | ||| |  | | | ||||||||   |  | ||     | ||||||||||||||||||||
Db     951 TGTTACCTTTGAAGACAACGGCACTTTGGTATTGGATCAAAACATCAACCAAGGCGCGGG 1010

Qy    1221 cggtttgtattttgaaggtgattttacggtctcgcctgaaaacaacga---aacgtggca 1277
           |||| ||| |||  |||| ||||  || |||     || ||| || ||    || ||||
Db    1011 CGGTCTGTTTTTCAAAGGCGATTACACAGTCAAAGGTGCAAATAATGACATCACTTGGCT 1070

Qy    1278 aggcgcgggcgttcatatcagtgaagacagtaccgttacttggaaagtaaacggcgtggc 1337
           ||| |||||  || || |    || | ||  |  ||   |||| |||| ||      |
Db    1071 AGGTGCGGGGATTGATGTTGCCGACGGCAAAAAAGTCGTTTGGCAAGTCAAAAATCCGAA 1130

Qy    1338 aaacgaccgcctgtccaaaatcggcaaaggcacgctgcacgttcaagccaaaggggaaaa 1397
           ||| |  || | ||||||||||||||||  |  |  |  | | || ||| |  ||
Db    1131 TGGTGACAGATTGGCAAAAATCGGCAAAGGCACATTAGAAATAAACGGCACAGGCGTTAA 1190

Qy    1398 ccaaggctcgatcagcgtgggcgacggtacagtcattttggatcagcaggcagacgataa 1457
           ||||||     | |  || |||||||| ||||| ||| || ||||  | || ||    ||
Db    1191 CCAAGGGCAATTAAAAGTCGGCGACGGCACAGTTATTCTGAATCAAAAAGCCGATTCAAA 1250

Qy    1458 aggcaaaaaacaagcctttagtgaaatcggcttgntcagcggcaggggtacggtgcaact 1517
           |||   || || ||     || ||||| |  |||||||  |  ||||  ||  | |
Db    1251 CCAAAAAGTCCAGGCTTTCTCCCAAGTCGGCATTGTCAGCGGACGCCGTACATTGGTATT 1310

Qy    1518 gaatgccgataatcagttcaacccccgacaaactctatttcggctttcgcggcggacgttt 1577
           |||      |||||| | || ||||| || || |||||||| || || ||||| |||||
Db    1311 AAATAGTTCAAATCAGATTAATCCCGATAACCTATATTTCGGTTTCCGTGGCGGTCGTTT 1370

Qy    1578 ggatttaaacgggcattcgctttcgttccaccgtattcaaaataccgatgaaggggcgat 1637
           ||||   || ||  ||    |  | ||  | |  || |  ||    ||||||||  |||
Db    1371 GGATGCCAATGGCAATGACTTGACTTTTGAACACATCCGCAACGTGGATGAAGGCGCGCG 1430

Qy    1638 gattgncnatcataatgccacaacaacatccaccgttaccattacagggaatgaaagtat 1697
           ||||  |  | || ||  |         | |||||  | ||  | || || ||   |  ||
Db    1431 CATTGTCAACCACAACACAGGCCACGCCTCCACAATCACATTAACCGGCAAATCATTAAT 1490

Qy    1698 tac 1700
           |||
Db    1491 TAC 1493


Search completed: June 30, 2002, 12:42:41
Job time: 69979 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2002, 02:30:47 ; Search time 415.2 Seconds
(without alignments)
528.893 Million cell updates/sec

Title: US-09-303-518D-571
Perfect score: 894
Sequence: 1 atgtttcgtttacaattcag.....accgctataaaacgccgtaa 894

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	38.6	4.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	2	37	4.1	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	3	35.8	4.0	7218	1	US-08-232-463-14	Sequence 14, Appl
c	4	35.2	3.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	5	34.2	3.8	289	4	US-09-007-005-17	Sequence 17, Appl
	6	34.2	3.8	289	4	US-09-244-796-17	Sequence 17, Appl
	7	34.2	3.8	2277	1	US-08-676-967-5	Sequence 5, Appli
	8	34.2	3.8	2277	1	US-08-676-974-5	Sequence 5, Appli
	9	34.2	3.8	2277	2	US-09-098-487-5	Sequence 5, Appli
	10	34	3.8	1198	3	US-09-248-335-27	Sequence 27, Appl
	11	33.8	3.8	2085	1	US-08-553-999B-1	Sequence 1, Appli
	12	33.8	3.8	3358	3	US-09-248-571-2	Sequence 2, Appli
c	13	33.6	3.8	2191	3	US-09-039-555B-12	Sequence 12, Appl
	14	33.6	3.8	35081	2	US-08-752-760A-1	Sequence 1, Appli
	15	33.4	3.7	2070	1	US-07-766-351-6	Sequence 6, Appli
	16	33.4	3.7	2070	1	US-08-059-032-6	Sequence 6, Appli
	17	33.4	3.7	2070	5	PCT-US91-07290-6	Sequence 6, Appli
	18	33.2	3.7	1155	4	US-09-422-487-4	Sequence 4, Appli
	19	33.2	3.7	1158	3	US-08-844-188-45	Sequence 45, Appl
	20	33.2	3.7	20235	1	US-07-642-734C-3	Sequence 3, Appli
	21	33.2	3.7	20235	3	US-08-439-009A-3	Sequence 3, Appli
	22	33	3.7	1119	4	US-08-987-943-2	Sequence 2, Appli
	23	33	3.7	1134	3	US-09-248-335-29	Sequence 29, Appl
	24	33	3.7	1755	4	US-08-987-943-1	Sequence 1, Appli
	25	33	3.7	7898	4	US-08-984-709A-49	Sequence 49, Appl
	26	32.8	3.7	1473	2	US-08-541-033A-25	Sequence 25, Appl
	27	32.8	3.7	1473	2	US-08-828-451-25	Sequence 25, Appl
	28	32.8	3.7	1506	2	US-08-541-033A-23	Sequence 23, Appl
	29	32.8	3.7	1506	2	US-08-828-451-23	Sequence 23, Appl
	30	32.8	3.7	1969	2	US-08-541-033A-7	Sequence 7, Appli
	31	32.8	3.7	1969	2	US-08-828-451-7	Sequence 7, Appli
	32	32.8	3.7	2096	2	US-08-541-033A-19	Sequence 19, Appl
	33	32.8	3.7	2096	2	US-08-828-451-19	Sequence 19, Appl
	34	32.8	3.7	2099	2	US-08-541-033A-3	Sequence 3, Appli
	35	32.8	3.7	2099	2	US-08-828-451-3	Sequence 3, Appli
	36	32.8	3.7	2137	2	US-08-541-033A-18	Sequence 18, Appl
	37	32.8	3.7	2137	2	US-08-828-451-18	Sequence 18, Appl
	38	32.8	3.7	2140	2	US-08-541-033A-1	Sequence 1, Appli
	39	32.8	3.7	2140	2	US-08-828-451-1	Sequence 1, Appli
	40	32.8	3.7	15872	4	US-09-105-537-1	Sequence 1, Appli
	41	32.8	3.7	43280	2	US-08-804-227C-1	Sequence 1, Appli
	42	32.6	3.6	520	2	US-08-997-080-179	Sequence 179, App
	43	32.6	3.6	520	2	US-08-997-362-179	Sequence 179, App
	44	32.6	3.6	520	4	US-09-095-855-179	Sequence 179, App
	45	32.6	3.6	520	4	US-09-324-542-179	Sequence 179, App
c	46	32.6	3.6	1269	1	US-08-844-064-5	Sequence 5, Appli
c	47	32.6	3.6	1269	3	US-09-009-433-5	Sequence 5, Appli
c	48	32.6	3.6	2652	2	US-08-953-492-1	Sequence 1, Appli
	49	32.4	3.6	423	1	US-08-470-179-111	Sequence 111, App
	50	32.4	3.6	476	3	US-08-050-259B-19	Sequence 19, Appl
c	51	32	3.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	52	31.8	3.6	3771	1	US-08-185-432-3	Sequence 3, Appli
	53	31.8	3.6	3771	1	US-08-185-432-23	Sequence 23, Appl
c	54	31.8	3.6	5063	1	US-08-185-432-1	Sequence 1, Appli
	55	31.6	3.5	2400	4	US-08-930-001-1	Sequence 1, Appli
	56	31.4	3.5	507	4	US-09-489-292-1	Sequence 1, Appli
	57	31.4	3.5	570	4	US-09-095-855-202	Sequence 202, App
c	58	31.4	3.5	5247	1	US-08-920-812-15	Sequence 15, Appl
c	59	31.4	3.5	5247	1	US-08-920-827-15	Sequence 15, Appl
c	60	31.4	3.5	5247	1	US-08-921-177-15	Sequence 15, Appl
c	61	31.4	3.5	5247	1	US-08-362-577C-15	Sequence 15, Appl
c	62	31.4	3.5	5247	2	US-08-920-828-15	Sequence 15, Appl
	63	31.4	3.5	5577	1	US-08-326-117B-1	Sequence 1, Appli
	64	31.4	3.5	5577	3	US-08-982-129-1	Sequence 1, Appli
	65	31.2	3.5	1668	4	US-09-385-028-20	Sequence 20, Appl
c	66	31.2	3.5	2256	1	US-07-794-393-1	Sequence 1, Appli
c	67	31.2	3.5	2256	1	US-08-001-711-1	Sequence 1, Appli
	68	31.2	3.5	11604	4	US-09-385-028-13	Sequence 13, Appl
	69	31.2	3.5	15079	4	US-09-385-028-1	Sequence 1, Appli
c	70	31	3.5	289	4	US-09-007-005-17	Sequence 17, Appl
c	71	31	3.5	289	4	US-09-244-796-17	Sequence 17, Appl
	72	31	3.5	1879	2	US-08-403-852D-5	Sequence 5, Appli
	73	31	3.5	1879	3	US-08-510-646B-5	Sequence 5, Appli
	74	31	3.5	1879	4	US-09-231-818-5	Sequence 5, Appli
	75	30.8	3.4	1483	4	US-09-262-749-1	Sequence 1, Appli
	76	30.8	3.4	1553	3	US-09-022-669-1	Sequence 1, Appli
	77	30.8	3.4	4800	3	US-08-941-445A-4	Sequence 4, Appli
	78	30.6	3.4	1155	2	US-08-387-942C-18	Sequence 18, Appl
c	79	30.6	3.4	1619	2	US-08-991-946A-2	Sequence 2, Appli
	80	30.6	3.4	2082	4	US-09-440-325A-2	Sequence 2, Appli
	81	30.6	3.4	3314	1	US-07-973-324A-5	Sequence 5, Appli
	82	30.6	3.4	3314	1	US-08-343-380-5	Sequence 5, Appli
	83	30.6	3.4	3314	4	US-09-072-435-5	Sequence 5, Appli
	84	30.6	3.4	3314	4	US-09-072-917A-5	Sequence 5, Appli
	85	30.6	3.4	12588	2	US-08-387-942C-1	Sequence 1, Appli
	86	30.6	3.4	28958	1	US-08-258-261B-6	Sequence 6, Appli
	87	30.6	3.4	28958	1	US-08-456-837-6	Sequence 6, Appli
	88	30.6	3.4	28958	1	US-08-457-342-6	Sequence 6, Appli
	89	30.6	3.4	28958	1	US-08-457-646A-6	Sequence 6, Appli
	90	30.6	3.4	28958	1	US-08-458-076A-6	Sequence 6, Appli
	91	30.6	3.4	28958	1	US-08-764-233A-4	Sequence 4, Appli
	92	30.6	3.4	28958	1	US-08-457-335A-6	Sequence 6, Appli
	93	30.6	3.4	28958	1	US-08-729-214-6	Sequence 6, Appli
	94	30.6	3.4	28958	3	US-09-028-934-6	Sequence 6, Appli
	95	30.6	3.4	49377	1	US-08-764-233A-1	Sequence 1, Appli
	96	30.4	3.4	2109	2	US-08-617-801A-5	Sequence 5, Appli
c	97	30.4	3.4	2181	4	US-09-158-767-10	Sequence 10, Appl
	98	30.4	3.4	2220	2	US-08-617-801A-1	Sequence 1, Appli
c	99	30.4	3.4	43280	2	US-08-804-227C-1	Sequence 1, Appli
c	100	30.2	3.4	1249	4	US-09-333-208-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             4.3%;  Score 38.6;  DB 4;  Length 4403765;
  Best Local Similarity   46.7%;  Pred. No. 2.6;
  Matches  122;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy     511 ggcaaaaccgcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcg 570
           | |||  ||| ||| | |    |    | ||||||      |  | |||| |||   | |
Db  361666 gccaacgccgggccgaacatgctggccgcggtcaacgcccccgcccaggcgctgttgggg 361725

Qy     571 ggcgaggcaaccatcatcctgcccgaccacgtcccttctccgcaggaaggcggcggcgtg 630
             ||       | |  || |    || ||   |  ||      |   | | | ||||||||
Db  361726 cgcccactgatcggcaacggtgccaacggggcgccgggcaccgggcaggccggcggcgac 361785

Qy     631 tgggcggatttttttcggcaaacctgcatacaccatgacactggcggcaaaattggcacac 690
           | | |   || ||||||||     |   ||  |  | |    ||  |    |  |||   |
Db  361786 ggtgggctgttgttcggcaacggcggcaacggcgggtccggcgcacccggacaggccggc 361845

Qy     691 gtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcccgacggacaaggcttc 750
           |     |||| |   |||  ||| ||| ||  ||   ||      ||  |     ||   |
Db  361846 ggggccggcggggcgaccgggttcttcggcaacggtggcaacggcggggacggcggggcc 361905

Qy     751 gtgttgcacatccgccccgtc 771
           |    | ||  | || ||| |
Db  361906 ggagcgaacggcggcgccggc 361926


RESULT   2
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


  Query Match             4.1%;  Score 37;  DB 4;  Length 4411529;
  Best Local Similarity   46.4%;  Pred. No. 7.2;
  Matches  121;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy     511 ggcaaaaccgcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcg 570
           | |||  ||| ||| | |    |    | ||||||      |  | |||| |||   | |
Db  361608 gccaacgccgggccgaacatgctggccgcggtcaacgcccccgcccaggcgctgttgggg 361667

Qy     571 ggcgaggcaaccatcatcctgcccgaccacgtcccttctccgcaggaaggcggcggcgtg 630
             ||       | |  || |    || ||   |  ||      |   | | | ||||||||
Db  361668 cgcccactgatcggcaacggtgccaacggggcgccgggcaccgggcaggccggcggcgac 361727

Qy     631 tgggcggatttttttcggcaaacctgcatacaccatgacactggcggcaaaattggcacac 690
           | | |   || ||||||||     |   ||  |  | |    ||  |    |  |||   |
Db  361728 ggtgggctgttgttcggcaacggcggcaacggcgggtccggcgcacccggacaggccggc 361787

Qy     691 gtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcccgacggacaaggcttc 750
           |     |||| |    ||  ||| ||| ||  ||   ||      ||  |     ||   |
Db  361788 ggggccggcggggcggccgggttcttcggcaacggtggcaacggcggggacggcggggcc 361847

Qy     751 gtgttgcacatccgccccgtc 771
           |    | ||  | || ||| |
Db  361848 ggagcgaacggcggcgccggc 361868


RESULT   3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             4.0%;  Score 35.8;  DB 1;  Length 7218;
  Best Local Similarity   5.5%;  Pred. No. 1;
  Matches   22;  Conservative  202;  Mismatches  179;  Indels    0;  Gaps    0;

Qy     150 aaaggaagaccgcgcgcgcatcgtcgccaatatgcggcaggcgggtttgaaccccgacac 209
           :::::::::  : : : : :  :  :  :: : : :: ::: :::   :::    :: : :
Db    1426 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1367

Qy     210 gcagacggtcaaagccgtttttgcggaaacggcaaaatgcggtttggaacttgcccccgc 269
           : ::: ::  ::::  :      : ::::: :: :::: : ::   ::::   :     :
Db    1366 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1307

Qy     270 gtttttcaaaaaaccggaagacatcgaaacaatgttcaaagcggtacacggctgggaaca 329
           :      ::::::  :::::: :  :::: :: :   :::: :: : : ::  ::::: :
Db    1306 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1247

Qy     330 cgtgcagcaggctttggacaagggcgaagggctgctgttcatcacgccgcacatcggcag 389
           : : :: :::    ::: ::::: ::::::  :  :   :  : :  : :: :  :: ::
Db    1246 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1187

Qy     390 ctacgatttgggcggacgctacatcagccagcagcttccgttccacctgaccgccatgta 449
             : ::   ::: ::: :  : :  ::  :: ::     :    :   :: :  : : :
Db    1186 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1127

Qy     450 caagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaa 509
           :::  :  ::::  :::: :: ::: ::::  : : ::: ::: :::: : : :: ::
Db    1126 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1067

Qy     510 aggcaaaaccgcgcccaccggcatacaaggggtcaaacaaatc 552
           |    || |  |  | ||| |||  ||||     ||  || ||
Db    1066 ATCGCAAGCTCCCTCGACCTGCAGCCAAGCTCGGAATTAATTC 1024


RESULT   4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


  Query Match             3.9%;  Score 35.2;  DB 4;  Length 4411529;
  Best Local Similarity  47.3%;  Pred. No. 21;
  Matches  107;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy     366 gttcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagct 425
           || |  |||  |   | | ||||         || | | || || |  ||||  || ||

Db 2980438 GTACGCCACCACCGCCGTGGGCACACCGACCATGCGGGCACACTCCTACAGCTTGCTGCG 2980379

Qy     426 tccgttccacctgaccgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcat 485
             |  ||   | |||||||  ||  | |    | ||     |           |   |||
Db 2980378 CGCAATCATGCAGACCGCCTTGGCCGACGACCTGATCGACTCCAACCCCTGCCGCATCTC 2980319

Qy     486 gcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaa 545
               |||  ||  |  ||| ||     ||| | |  |||| ||  | |  | | | |  |
Db 2980318 AGGCGCGTCCACCGCCCGCCGCGTCCACAAGATCAGGCCCGCCACCCTCGACGAGCTGGA 2980259

Qy     546 acaaatcatcaaggccctgcgcgcgggcgaggcaaccatcatcctg 591
           |   |||| ||| ||| ||| ||    | |     |  || | |||
Db 2980258 AACCATCACCAAAGCCATGCCCGACCCCTACCAGGCGTTCGTGCTG 2980213


RESULT   5
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17


  Query Match             3.8%;  Score 34.2;  DB 4;  Length 289;
  Best Local Similarity   2.2%;  Pred. No. 0.7;
  Matches    4;  Conservative   86;  Mismatches   91;  Indels    0;  Gaps    0;

Qy     459 gaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaac 518
           |: :  :: :  :: : ::: :  :: : ::: : ::: :  :    ::: : ::: :
Db      56 grnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnr 115

Qy     519 cgcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcgggcgaggc 578
           :: :  ::   ::: : : : :  :: : : :    :: :  :  : ::: : ::: :
Db     116 srnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnr 175

Qy     579 aaccatcatcctgcccgaccacgtcccttctccgcaggaaggcggcggcgtgtgggcgga 638
           :    ::    :    ::    :        : ::: : :::   : : :::|:|:|:
Db     176 srnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgr 235

Qy     639 t 639
           :
Db     236 u 236


RESULT   6
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
```

```
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17


  Query Match             3.8%;  Score 34.2;  DB 4;  Length 289;
  Best Local Similarity   2.2%;  Pred. No. 0.7;
  Matches    4;  Conservative    86;  Mismatches    91;  Indels    0;  Gaps    0;

Qy     459 gaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaac 518
           |: :  :: :  :: : ::: :  :: : ::: : ::: :  :    ::: : ::: :
Db      56 grnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnr 115

Qy     519 cgcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcgggcgaggc 578
           :: :  ::   ::: : : : :  :: : : :    :: :  :  : ::: : ::: :
Db     116 srnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnr 175

Qy     579 aaccatcatcctgcccgaccacgtcccttctccgcaggaaggcggcggcgtgtgggcgga 638
           :    ::    :    ::     :        : ::: : :::   : : :::|:|:|:
Db     176 srnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgr 235

Qy     639 t 639
           :
Db     236 u 236


RESULT   7
US-08-676-967-5
; Sequence 5, Application US/08676967
; Patent No. 5747317
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,967
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-967-5


  Query Match             3.8%;  Score 34.2;  DB 1;  Length 2277;
  Best Local Similarity   44.4%;  Pred. No. 1.8;
  Matches  138;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy     306 caaagcggtacacggctgggaacacgtgcagcaggctttggacaagggcgaagggctgct 365
           ||| || |   | |||  |     ||   |   |||   | | || ||   | ||||||
Db    1356 CAAGGCCGCCGAGGGCGTGAGCGCCGCCGACATGGCCAAGCGCGAGCGCTTCGAGCTGCT 1415

Qy     366 gttcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagct 425
           |    |  |  ||  |  | |    ||   || |   | |||  |  |     | ||
Db    1416 GAAGCACCAGAAGCTGAAGGACCAGAACATCTTCGTGAGCCGCACCCGCCTGTGCCTGCA 1475

Qy     426 tccgttccacctgaccgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcat 485
             | | |  | |||      ||||||  || |   |       | |   |   | ||
Db    1476 CAACCTGCCCAAGGCCGTGGACGACAAGCAGCTGCGCAAGCTGCTGCTGAGCGCCACCAG 1535

Qy     486 gcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaa 545
             | |    | ||||||  ||| |       || |     || | |  | || ||  |
Db    1536 CGGCGAGAAGGGCGTGCGCATCAAGGAGTGCCGCGTGATGCGCGACCTGAAGGGCGTGCA 1595

Qy     546 acaaatcatcaaggccctgcgcgcgggcgaggcaaccatcatcctgcccgaccacgtccc 605
             | ||| |||| || | ||  |||| | ||   |  |    | |  || ||||  |
Db    1596 CGGCAACATGAAGGGCCAGAGCCTGGGCTACGCCTTCGCCGAGTTCCAGGAGCACGAGCA 1655

Qy     606 ttctccgcagg 616
             | | | |||
Db    1656 CGCCCTGAAGG 1666


RESULT   8
US-08-676-974-5
; Sequence 5, Application US/08676974
; Patent No. 5770422
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,974
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
```

```
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-974-5


  Query Match             3.8%;  Score 34.2;  DB 1;  Length 2277;
  Best Local Similarity   44.4%;  Pred. No. 1.8;
  Matches  138;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy     306 caaagcggtacacggctgggaacacgtgcagcaggctttggacaagggcgaagggctgct 365
           ||| || |   | |||  |     ||   |   |||   |  | || ||   | ||||||
Db    1356 CAAGGCCGCCGAGGGCGTGAGCGCCGCCGACATGGCCAAGCGCGAGCGCTTCGAGCTGCT 1415

Qy     366 gttcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagct 425
           |     |  |  ||  |  | |    ||   || |   | |||  |  |      | ||
Db    1416 GAAGCACCAGAAGCTGAAGGACCAGAACATCTTCGTGAGCCGCACCCGCCTGTGCCTGCA 1475

Qy     426 tccgttccacctgaccgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcat 485
               | | |  | |||      ||||| || |   |       | |   |   | ||
Db    1476 CAACCTGCCCAAGGCCGTGGACGACAAGCAGCTGCGCAAGCTGCTGCTGAGCGCCACCAG 1535

Qy     486 gcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaa 545
           | |    | ||||||  ||| |       || |     || | |  | || ||  |
Db    1536 CGGCGAGAAGGGCGTGCGCATCAAGGAGTGCCGCGTGATGCGCGACCTGAAGGGCGTGCA 1595

Qy     546 acaaatcatcaaggccctgcgcgcgggcgaggcaaccatcatcctgcccgaccacgtccc 605
             | ||| |||| || | ||  |||| | ||   |  |    | |  || ||||  |
Db    1596 CGGCAACATGAAGGGCCAGAGCCTGGGCTACGCCTTCGCCGAGTTCCAGGAGCACGAGCA 1655

Qy     606 ttctccgcagg 616
             | | | |||
Db    1656 CGCCCTGAAGG 1666


RESULT  9
US-09-098-487-5
; Sequence 5, Application US/09098487
; Patent No. 5917025
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, Kathleen
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,487
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-098-487-5


  Query Match             3.8%;  Score 34.2;  DB 2;  Length 2277;
  Best Local Similarity   44.4%;  Pred. No. 1.8;
  Matches  138;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy     306 caaagcggtacacggctgggaacacgtgcagcaggctttggacaagggcgaagggctgct 365
           ||| || |   | |||  |     ||   |   |||   |  | || ||   | ||||||
Db    1356 CAAGGCCGCCGAGGGCGTGAGCGCCGCCGACATGGCCAAGCGCGAGCGCTTCGAGCTGCT 1415

Qy     366 gttcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagct 425
           |     |  |  ||  |  | |    ||   || |   | |||  |  |      | ||
Db    1416 GAAGCACCAGAAGCTGAAGGACCAGAACATCTTCGTGAGCCGCACCCGCCTGTGCCTGCA 1475

Qy     426 tccgttccacctgaccgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcat 485
               | | |  | |||      ||||| || |   |       | |   |   | ||
Db    1476 CAACCTGCCCAAGGCCGTGGACGACAAGCAGCTGCGCAAGCTGCTGCTGAGCGCCACCAG 1535

Qy     486 gcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaa 545
           | |    | ||||||  ||| |       || |     || | |  | || ||  |
Db    1536 CGGCGAGAAGGGCGTGCGCATCAAGGAGTGCCGCGTGATGCGCGACCTGAAGGGCGTGCA 1595

Qy     546 acaaatcatcaaggccctgcgcgcgggcgaggcaaccatcatcctgcccgaccacgtccc 605
             | ||| |||| || | ||  |||| | ||   |  |    | |  || ||||  |
Db    1596 CGGCAACATGAAGGGCCAGAGCCTGGGCTACGCCTTCGCCGAGTTCCAGGAGCACGAGCA 1655

Qy     606 ttctccgcagg 616
             | | | |||
Db    1656 CGCCCTGAAGG 1666


RESULT  10
US-09-248-335-27
; Sequence 27, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
;  APPLICANT: MCGONIGLE, BRIAN
;  APPLICANT: O'KEEF, DANIEL
;  TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE: CL-1128-A
;  CURRENT APPLICATION NUMBER: US/09/248,335
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: 08/924,759
;  EARLIER FILING DATE: 1997-September-05
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 27
;   LENGTH: 1198
;   TYPE: DNA
;   ORGANISM: maize
US-09-248-335-27


  Query Match             3.8%;  Score 34;  DB 3;  Length 1198;
  Best Local Similarity   46.0%;  Pred. No. 1.5;
  Matches  115;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy     344 tggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcg 403
           ||| |     ||| |   | || | ||  |    | || ||| | || |||| ||
Db     546 tgggcgtctacgacgacatcctctccaagaacgagtacctcgccggcgacgacttcacac 605

Qy     404 gacgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaa 463
             | || ||  ||| | ||     | ||||   | || ||  | |        |   |
```

```
Db    606 tggccgacctctcccacctgccaaactcccactacatcgtcaactcctccgacaggggca 665

Qy    464 tcaaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgc 523
          ||     |   |   | | ||| | || ||||  |  || |||   |   |||
Db    666 ggaagctcttcaccgccaggaagcacgtggccaggtggtacgacaagatctccacccgcg 725

Qy    524 ccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcgggcgaggcaacca 583
          | || | |  || |  | ||       |   ||  ||  |||||  |||| | |  |
Db    726 actcctggaggcaggtcatgaagatgcagagggagcaccccggcgcgttcgagtgatgcg 785

Qy    584 tcatcctgcc 593
          || | || ||
Db    786 tcgtgcttcc 795


RESULT 11
US-08-553-999B-1
; Sequence 1, Application US/08553999B
; Patent No. 5773599
;  GENERAL INFORMATION:
;    APPLICANT:  Govindaswamy Chinnadurai
;    TITLE OF INVENTION:  Isolation and characterization of cDNA for E1A C-terminal bind
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  G. Chinnadurai
;      STREET:  3681 Park Avenue
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch; 713 Kb Storage
;      COMPUTER:  IBM Compatible (Gateway 2000)
;      OPERATING SYSTEM:  Windows 3.1
;      SOFTWARE:  Word for Windows (6.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/553,999B
;      FILING DATE:  11/06/95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5773599e
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2085 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  human
;      INDIVIDUAL ISOLATE:  pACT30
;      CELL TYPE:  B-cell
;    IMMEDIATE SOURCE:
;      LIBRARY:  cDNA
;      CLONE:  pACT30
;    PUBLICATION INFORMATION:
;      AUTHORS:  Ute Schaeper
;      AUTHORS:  Janice M. Boyd
;      AUTHORS:  Sulekha Verma
;      AUTHORS:  Erik Uhlmann
;      AUTHORS:  T. Subramanian
;      AUTHORS:  G. Chinnadurai
;      TITLE:  Molecular cloning and characterization of a cellular phosphoprotein that
;      JOURNAL:  Proceedings of the National Academy of Sciences United States of Americ
;      VOLUME:  92
;      ISSUE:  23
;      PAGES:  10667-10671
;      RELEVANT RESIDUES IN SEQ ID NO:  1:
US-08-553-999B-1


  Query Match             3.8%;  Score 33.8;  DB 1;  Length 2085;
  Best Local Similarity   47.8%;  Pred. No. 2.2;
  Matches   98;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy    435 cctgaccgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcggg 494
          | ||  | ||     |  ||  ||  |  |||| | |   | | | |     |   ||||
Db   1135 CATGGACCCCGCCGTCGTGCACCCTGAGCTCAATGGGGCTGCCTATAGGTACCCTCCGGG 1194

Qy    495 cagggtgcgcggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaaacaaatcat 554
          |  |||| |||        ||    |   ||  | |   |  ||||  ||   |  | |
Db   1195 CGTGGTGGGCGTGGCCCCCACTGGCATCCCAGCTGCTGTGGAAGGTATCGTCCCCAGCGC 1254

Qy    555 caaggccctgcgcgcgggcgaggcaaccatcatcctgcccgaccacgtcccttctccgca 614
          || | ||||| |   ||| | | |  |  |   ||  ||   ||||| |||||||||
Db   1255 CATGTCCCTGTCCCACGGCCTGCCCCCTGTGGCCCACCCGCCCCACGCCCCTTCTCCTGG 1314

Qy    615 ggaaggcggcggcgtgtgggcggat 639
            ||  || |        |||||||
Db   1315 CCAAACCGTCAAGCCCGAGGCGGAT 1339


RESULT 12
US-09-248-571-2
; Sequence 2, Application US/09248571
; Patent No. 6136539
; GENERAL INFORMATION:
;  APPLICANT: BASBAUM, CAROL
;  APPLICANT: GALLUP, MARIANNE
;  APPLICANT: DAIZONG, LI
;  APPLICANT: GEBREMICHAEL, ASSEFA
;  APPLICANT: GENSCH, ERIN
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITION OF MUC-5 MUCIN
;  TITLE OF INVENTION: GENE EXPRESSION
;  FILE REFERENCE: UCSF12/02
;  CURRENT APPLICATION NUMBER: US/09/248,571
;  CURRENT FILING DATE: 1999-02-11
;  EARLIER APPLICATION NUMBER: 60/074,398
;  EARLIER FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3358
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-248-571-2


  Query Match             3.8%;  Score 33.8;  DB 3;  Length 3358;
  Best Local Similarity   58.4%;  Pred. No. 2.7;
  Matches   59;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy    345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
          |||| || || | | | |||  | || ||  | ||| || |||| ||||    |||  |
Db   3002 ggacgagagccaggaggtgccatacaccatccggcagatgggcatctacctggtggtgga 3061

Qy    405 acgctacatcagccagcagcttccgttccacctgaccgcca 445
            | ||||  ||| |  ||| | ||   ||  ||   |||
Db   3062 caccgacattggcctggtgctgctgtgggacaagaagacca 3102


RESULT 13
US-09-039-555B-12/c
; Sequence 12, Application US/09039555B
; Patent No. 6033856
;  GENERAL INFORMATION:
;    APPLICANT:  Koerner, Kathrin
;    APPLICANT:  Mueller, Rolf
;    APPLICANT:  Sadlacek, Hans-Harald
;    TITLE OF INVENTION:  PROMOTER OF THE CDC25B GENE, ITS
;    TITLE OF INVENTION:  PREPARATION AND USE
```

```
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/039,555B
;      FILING DATE:  16-MAR-1998
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19710643.9
;      FILING DATE:  14-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  016779/0131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2191 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-039-555B-12


  Query Match             3.8%;  Score 33.6;  DB 3;  Length 2191;
  Best Local Similarity   46.6%;  Pred. No. 2.6;
  Matches  108;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy     390 ctacgatttgggcggacgctacatcagccagcagcttccgttccacctgaccgccatgta 449
           || | |    ||| | || |||  | ||     |   ||    |  | |    |   | |
Db     236 CTCCCACAGCGGCCGCCGGTACCACTGCTCCTCGAAGCCCCGGCGTCGGTTGTCAGAGAA 177

Qy     450 caagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaa 509
             | ||| |||  || |   |     ||    | ||||  |  ||   | || |  |
Db     176 GTCGGCGCGGAAGCTCCAGAGGCCGTCCAGCTCCTTGCACTCCCGCGACGGGCTCTCCTG 117

Qy     510 aggcaaaaccgcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgc 569
             ||  | | |   ||  || |||  |  | |   | |    || ||| | || | ||||
Db     116 GGGGTACAGCATCCCGCCCTGCAGCCCCAGCGCGCAGCCCCACAACAACGGCCCGAGCGC 57

Qy     570 gggcgaggcaaccatcatcctgcccgaccacgtcccttctccgcaggaaggc 621
           | | ||||||||  |  ||  |  | |    | ||   ||| | |   |||
Db      56 CGCCCAGGCAACCGCCGACCCCCGGGCCATGCTTCCCGGTCCCCCGCTCGGC 5


RESULT  14
US-08-752-760A-1
; Sequence 1, Application US/08752760A
; Patent No. 5877011
;  GENERAL INFORMATION:
;    APPLICANT:  Armentano, Donna
;    APPLICANT:  Gregory, Richard J.
;    APPLICANT:  Smith, Alan E.
;    TITLE OF INVENTION:  CHIMERIC ADENOVIRAL VECTORS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Baker & Botts, L.L.P.
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  U.S.A.
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/752,760A
;      FILING DATE:  20-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seide, Rochelle K
;      REGISTRATION NUMBER:  32,300
;      REFERENCE/DOCKET NUMBER:  A31385
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-705-5000
;      TELEFAX:  212-705-5020
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35081 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-752-760A-1


  Query Match             3.8%;  Score 33.6;  DB 2;  Length 35081;
  Best Local Similarity   50.6%;  Pred. No. 8.9;
  Matches   81;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     695 aaggcgtgaaaaccctgtttttctgctgcgaacgcctgcccgacggacaaggcttcgtgt 754
           ||| |   ||   ||   |  || ||  ||| | | |||||  | |        | | |
Db   14713 AAGTCAGCAACTACCCCGTGGTCGGCGCCGAGCTCATGCCCTTCCGCGCCAAGAGCTTTT 14772

Qy     755 tgcacatccgccccgtccaaggggaattgaacggcaacaaagcccacgatgccgccgtgt 814
             ||  || | ||||| |      |  | | | ||| | |  ||  |    ||  ||| |
Db   14773 ACAACGACCTCGCCGTCTACTCCCAGCTCATCCGCAGCTACACCTCCCTCACCCACGTCT 14832

Qy     815 tcaaccgcaataccgaatattggatacgccgttttccgac 854
           ||||||||    ||||  |   ||| | | |    |||| |
Db   14833 TCAACCGCTTCCCCGACAACCAGATCCTCTGCCGTCCGCC 14872


RESULT  15
US-07-766-351-6
; Sequence 6, Application US/07766351
; Patent No. 5292652
;  GENERAL INFORMATION:
;    APPLICANT:  Sinha, Sukanto
;    APPLICANT:  Seubert, Peter A.
;    APPLICANT:  Dovey, Harry F.
;    APPLICANT:  McConlogue, Lisa C.
;    APPLICANT:  Little, Sheila P.
;    APPLICANT:  Johnstone, Edward M.
;    TITLE OF INVENTION:  Amyloidin Protease and Uses Thereof
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Athena Neurosciences, Inc.
;      STREET:  800F Gateway Blvd.
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/766,351
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy, Lisabeth Feix
;      REGISTRATION NUMBER:  31547
;      REFERENCE/DOCKET NUMBER:  17796-002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 877-0900
;      TELEFAX:  (415) 877-8370
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2070 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-07-766-351-6


  Query Match             3.7%;  Score 33.4;  DB 1;  Length 2070;
  Best Local Similarity   47.4%;  Pred. No. 2.9;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     381 catcggcagctacgatttgggcggacgctacatcagccagcagcttccgttccacctgac 440
           || | || | | | |   ||      ||   |||  | || ||||   |   |  | | |
Db     732 CAACTGCCGGTGCAAGGAGGAGAACTGCGCTATCCTCAAGGAGCTGGTGACGCTGCGGGC 791

Qy     441 cgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggt 500
           |   | || |   | || |    || |  ||   | | |      | | || | . |   |
Db     792 CCAGAAGTCCCGCCTGCTGGGGTTCCACACGCACGCCGACTATGTCCTGGAGATGAACAT 851

Qy     501 gcgcggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaaacaaatcatcaaggc 560
           |  |   |   || | |||| | |||||  | || | | |    || |   | ||| |
Db     852 GGCCAAGACCAGCCAGACCGTGGCCACCTTCCTAGATGAGCTGGCGCAGAAGCTGAAGCC 911

Qy     561 cctgcgcgcgggcgaggcaaccatcatcctg 591
           |||| | | |   |||    |  | || |||
Db     912 CCTGGGGGAGCAGGAGCGTGCGGTGATTCTG 942


RESULT  16
US-08-059-032-6
; Sequence 6, Application US/08059032
; Patent No. 5424205
;  GENERAL INFORMATION:
;    APPLICANT:  Sinha, Sukanto
;    APPLICANT:  Seubert, Peter A.
;    APPLICANT:  Dovey, Harry F.
;    APPLICANT:  McConlogue, Lisa C.
;    APPLICANT:  Little, Sheila P.
;    APPLICANT:  Johnstone, Edward M.
;    TITLE OF INVENTION:  Amyloidin Protease and Uses Thereof
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Street Tower,
;      STREET:  Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/059,032
;      FILING DATE:  19930507
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, William E.
;      REGISTRATION NUMBER:  30,223
;      REFERENCE/DOCKET NUMBER:  15270-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 326-2400
;      TELEFAX:  (415) 326-2422
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2070 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-059-032-6


  Query Match             3.7%;  Score 33.4;  DB 1;  Length 2070;
  Best Local Similarity   47.4%;  Pred. No. 2.9;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     381 catcggcagctacgatttgggcggacgctacatcagccagcagcttccgttccacctgac 440
           || | || | | | |   ||      ||   |||  | || ||||   |   |  | | |
Db     732 CAACTGCCGGTGCAAGGAGGAGAACTGCGCTATCCTCAAGGAGCTGGTGACGCTGCGGGC 791

Qy     441 cgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggt 500
           |   | || |   | || |    || |  ||   | | |      | | || | . |   |
Db     792 CCAGAAGTCCCGCCTGCTGGGGTTCCACACGCACGCCGACTATGTCCTGGAGATGAACAT 851

Qy     501 gcgcggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaaacaaatcatcaaggc 560
           |  |   |   || | |||| | |||||  | || | | |    || |   | ||| |
Db     852 GGCCAAGACCAGCCAGACCGTGGCCACCTTCCTAGATGAGCTGGCGCAGAAGCTGAAGCC 911

Qy     561 cctgcgcgcgggcgaggcaaccatcatcctg 591
           |||| | | |   |||    |  | || |||
Db     912 CCTGGGGGAGCAGGAGCGTGCGGTGATTCTG 942


RESULT  17
PCT-US91-07290-6
; Sequence 6, Application PC/TUS9107290
;  GENERAL INFORMATION:
;    APPLICANT:  Sinha, Sukanto
;    APPLICANT:  Seubert, Peter A.
;    APPLICANT:  Dovey, Harry F.
;    APPLICANT:  McConlogue, Lisa C.
;    APPLICANT:  Little, Sheila P.
;    APPLICANT:  Johnstone, Edward M.
;    TITLE OF INVENTION:  Amyloidin Protease and Uses Thereof
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Athena Neurosciences, Inc.
;      STREET:  800F Gateway Blvd.
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/07290
;      FILING DATE:  19911004
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy, Lisabeth Feix
;      REGISTRATION NUMBER:  31547
```

```
;      REFERENCE/DOCKET NUMBER:  17796-002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 877-0900
;      TELEFAX:  (415) 877-8370
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2070 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US91-07290-6


  Query Match             3.7%;  Score 33.4;  DB 5;  Length 2070;
  Best Local Similarity   47.4%;  Pred. No. 2.9;
  Matches  100;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy    381 catcggcagctacgatttgggcggacgctacatcagccagcagcttccgttccacctgac 440
          || | || | | | |   ||       ||   ||| | || ||||   |   |  | | |
Db    732 CAACTGCCGGTGCAAGGAGGAGAACTGCGCTATCCTCAAGGAGCTGGTGACGCTGCGGGC 791

Qy    441 cgccatgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggt 500
          |   | || |   | || |    || |  ||   | | |     | | || |   |   |
Db    792 CCAGAAGTCCCGCCTGCTGGGGTTCCACACGCACGCCGACTATGTCCTGGAGATGAACAT 851

Qy    501 gcgcggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaaacaaatcatcaaggc 560
          | |   |   || | |||| | ||||| | || | | |    || |   | ||| |
Db    852 GGCCAAGACCAGCCAGACCGTGGCCACCTTCCTAGATGAGCTGGCGCAGAAGCTGAAGCC 911

Qy    561 cctgcgcgcgggcgaggcaaccatcatcctg 591
          |||| | | |   |||    |  | || |||
Db    912 CCTGGGGGAGCAGGAGCGTGCGGTGATTCTG 942


RESULT  18
US-09-422-487-4
; Sequence 4, Application US/09422487
; Patent No. 6340593
; GENERAL INFORMATION:
;  APPLICANT:  Cardineau, Guy A.
;  APPLICANT:  Stelman, Steven J.
;  APPLICANT:  Narva, Kenneth E.
;  TITLE OF INVENTION:  Plant-Optimized Polynucleotides Encoding Approximately
;  TITLE OF INVENTION:  15 kDa and Approximately 45 kDa Pesticidal Proteins
;  FILE REFERENCE:  MA723/4X
;  CURRENT APPLICATION NUMBER: US/09/422,487
;  CURRENT FILING DATE:  1999-10-21
;  PRIOR APPLICATION NUMBER:  60/105,408
;  PRIOR FILING DATE:  1998-10-23
;  PRIOR APPLICATION NUMBER:  60/105,359
;  PRIOR FILING DATE:  1998-10-23
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 1155
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
;   OTHER INFORMATION: Bacillus thuringiensis toxin gene
US-09-422-487-4


  Query Match             3.7%;  Score 33.2;  DB 4;  Length 1155;
  Best Local Similarity   46.8%;  Pred. No. 2.5;
  Matches  104;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy    324 ggaacacgtgcagcaggctttggacaagggcgaagggctgctgttcatcacgccgcacat 383
          | || ||  | |   |    | | |||||||       ||    || || |||  || |
Db    618 gaaatacaagtactggaacctcgccaagggctccaacgtgtccctcctcccgcaccagaa 677

Qy    384 cggcagctacgatttgggcggacgctacatcagccagcagcttccgttccacctgaccgc 443
          ||||||||||  |   |   |  ||  |   |  | |||    |   |  | | | | |
Db    678 gcgcagctacgactacgagtggggcaccgagaagaaccagaagaccaccatcatcaacac 737

Qy    444 catgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtgcg 503
          | ||  |  || |   || ||| |  ||       ||  ||  |  | |||    ||   |
Db    738 cgtgggcctgcagatcaacatcgactcggggatgaagttcgaggtgccggaggtgggcgg 797

Qy    504 cggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaa 545
          |||||  |   | | |  | ||     ||   | | | | ||
Db    798 cggcaccgaggacatcaagacccagctcaccgaggagctgaa 839


RESULT  19
US-08-844-188-45
; Sequence 45, Application US/08844188
; Patent No. 6127180
;  GENERAL INFORMATION:
;    APPLICANT:  Narva, Kenneth E.
;    APPLICANT:  Schnepf, H. Ernest
;    APPLICANT:  Knuth, Mark
;    APPLICANT:  Pollard, Michael R.
;    APPLICANT:  Cardineau, Guy
;    APPLICANT:  Schwab, George E.
;    TITLE OF INVENTION:  Pesticidal Toxins
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/844,188
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/633,993
;      FILING DATE:  19-APR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sanders, Jay M.
;      REGISTRATION NUMBER:  39,355
;      REFERENCE/DOCKET NUMBER:  MA-703C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  352-375-8100
;      TELEFAX:  352-372-5800
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1158 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-844-188-45


  Query Match             3.7%;  Score 33.2;  DB 3;  Length 1158;
  Best Local Similarity   46.8%;  Pred. No. 2.5;
  Matches  104;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy    324 ggaacacgtgcagcaggctttggacaagggcgaagggctgctgttcatcacgccgcacat 383
          | || ||  | |   |    | | |||||||       ||    || || |||  || |
Db    618 GAAATACAAGTACTGGAACCTCGCCAAGGGCTCCAACGTGTCCCTCCTCCCGCACCAGAA 677

Qy    384 cggcagctacgatttgggcggacgctacatcagccagcagcttccgttccacctgaccgc 443
```

```
          ||||||||||| |  |   |  ||  |   |   | |||    |   |  | | | | |
Db    678 GCGCAGCTACGACTACGAGTGGGGCACCGAGAAGAACCAGAAGACCACCATCATCAACAC 737

Qy    444 catgtacaagccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtgcg 503
          | ||  |  || |   || ||| |  ||       ||  ||  |  | |||    ||   |
Db    738 CGTGGGCCTGCAGATCAACATCGACTCGGGGATGAAGTTCGAGGTGCCGGAGGTGGGCGG 797

Qy    504 cggcaaaggcaaaaccgcgcccaccggcatacaaggggtcaa 545
          |||||  |   | | |  | ||     ||   | | | | ||
Db    798 CGGCACCGAGGACATCAAGACCCAGCTCACCGAGGAGCTGAA 839


RESULT 20
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  97..1482
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1693..2670
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3406..3921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4171..4428
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..10722
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..5847
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6054..7026
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7165..9216
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9433..9984
;      OTHER INFORMATION:  /function= "approximate span
;      OTHER INFORMATION:  beta-ketoreductase of module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10225..10483
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 4"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10723..20235
;      OTHER INFORMATION:  /codon_start= 10723
;      OTHER INFORMATION:  /function= "gene =eryA"
;      OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;      OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10723..15165
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10831..12174
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  12379..13350
;      OTHER INFORMATION:  /function= "approximatr span of
;      OTHER INFORMATION:  acyltransferase domain of module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  14062..14610
```

```
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-07-642-734C-3


  Query Match             3.7%;  Score 33.2;  DB 1;  Length 20235;
  Best Local Similarity   55.1%;  Pred. No. 9.1;
  Matches   65;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     520 gcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcgggcgaggca 579
           ||  || ||||| |    ||||| |   |     |   ||| |||||   ||||||
Db   11851 GCCGCCGCCGGCGTCGCCGGGGTGATGAAGGCGGTGCTGGCGCTGCGGCACGGCGAGATG 11910

Qy     580 accatcatcctgcccgaccacgtcccttctccgcaggaaggcggcggcgtgtgggcgg 637
             |   ||  |||| |   | |||  || || ||||||   |   | | | || | ||||
Db   11911 CCGCGCACGCTGCACTTCGACGAGCCCTCGCCGCAGATCGAGTGGGACCTGGGCGCGG 11968


RESULT  21
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  97..1482
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1693..2670
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3406..3921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4171..4428
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..10722
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..5847
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6054..7026
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 4"
```

```
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-08-439-009A-3


  Query Match             3.7%;  Score 33.2;  DB 3;  Length 20235;
  Best Local Similarity   55.1%;  Pred. No. 9.1;
  Matches   65;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy    520 gcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcgggcgaggca 579
          ||  || ||||| |    ||||| |   |      |   ||| |||||    ||||||
Db  11851 GCCGCCGCCGGCGTCGCCGGGGTGATGAAGGCGGTGCTGGCGCTGCGGCACGGCGAGATG 11910

Qy    580 accatcatcctgcccgaccacgtcccttctccgcaggaaggcggcggcgtgtgggcgg 637
           |   ||  |||| |  | |||  || || ||||||   |   | | | || | ||||
Db  11911 CCGCGCACGCTGCACTTCGACGAGCCCTCGCCGCAGATCGAGTGGGACCTGGGCGCGG 11968


RESULT  22
US-08-987-943-2
; Sequence 2, Application US/08987943
; Patent No. 6210967
; GENERAL INFORMATION:
;  APPLICANT: Bard, Jonathan A.
;  TITLE OF INVENTION: DNA Encoding A Mammalian LPA Receptor and Uses Thereof
;  FILE REFERENCE: 55183
;  CURRENT APPLICATION NUMBER: US/08/987,943
;  CURRENT FILING DATE: 1997-12-10
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 2
;   LENGTH: 1119
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-987-943-2


  Query Match             3.7%;  Score 33;  DB 4;  Length 1119;
  Best Local Similarity   63.0%;  Pred. No. 2.9;
  Matches   51;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy     28 ccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaatgcctctccctg 87
          ||  || ||||  | |||||  || | ||| || || |||||  |   | |  |  | ||
Db    187 ccggctatgcgctcggccatcaacctgctgctggccaccctggccttctccgacatcatg 246

Qy     88 ctgtcgctttcctgtctgcac 108
          ||||| || | |||  ||| |
Db    247 ctgtccctctgctgcatgccc 267


RESULT  23
US-09-248-335-29
; Sequence 29, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
;  APPLICANT: MCGONIGLE, BRIAN
;  APPLICANT: O'KEEF, DANIEL
;  TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE: CL-1128-A
;  CURRENT APPLICATION NUMBER: US/09/248,335
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: 08/924,759
;  EARLIER FILING DATE: 1997-September-05
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 29
;   LENGTH: 1134
;   TYPE: DNA
;   ORGANISM: maize
US-09-248-335-29
```

```
  Query Match             3.7%;  Score 33;  DB 3;  Length 1134;
  Best Local Similarity   44.6%;  Pred. No. 2.9;
  Matches  129;  Conservative    0;  Mismatches  160;  Indels   0;  Gaps    0;

Qy    344 tggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcg 403
          |||  |     |||  |   | || | ||  |    | || ||| | || |||| ||   |
Db    503 tgggcgtctacgacgacatcctctccaagaacgagtacctcgccggcgacgacttcaccc 562

Qy    404 gacgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaa 463
              | || ||  ||| | ||    | ||||   | || ||  | |          |   |
Db    563 tggccgacctctcccacctgcccaactcccactacatcgtcaactcctccgacaggggca 622

Qy    464 tcaaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgc 523
            ||     |   |   | | ||| |  | ||||  |  || |||   |   |||
Db    623 gaaagctcttcaccgccaggaagcacgtcgccaggtggtacgacaagatctccacccgcg 682

Qy    524 ccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcgggcgaggcaacca 583
          | | | |  || |  ||||       |  ||  ||  ||||| |||| | |
Db    683 actcgtggaggcaggtcatcaagatgcagagggagcaccccggcgcgttcgagtgatcgg 742

Qy    584 tcatcctgcccgaccacgtcccttctccgcaggaaggcggcggcgtgtg 632
          ||      |   |   |  | |   ||| | | ||| ||  | |||
Db    743 tcggtggtctacgcggtgatgcatgcatgcatgcatgccgccggcgtg 791


RESULT  24
US-08-987-943-1
; Sequence 1, Application US/08987943
; Patent No. 6210967
; GENERAL INFORMATION:
;  APPLICANT: Bard, Jonathan A.
;  TITLE OF INVENTION: DNA Encoding A Mammalian LPA Receptor and Uses Thereof
;  FILE REFERENCE: 55183
;  CURRENT APPLICATION NUMBER: US/08/987,943
;  CURRENT FILING DATE: 1997-12-10
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 1
;   LENGTH: 1755
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-987-943-1


  Query Match             3.7%;  Score 33;  DB 4;  Length 1755;
  Best Local Similarity   63.0%;  Pred. No. 3.5;
  Matches   51;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy     28 ccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaatgcctctccctg 87
          ||  || ||||  | |||||  || | ||| || || |||||  |   | |  |  | ||
Db    536 ccggctatgcgctcggccatcaacctgctgctggccaccctggccttctccgacatcatg 595

Qy     88 ctgtcgctttcctgtctgcac 108
          ||||| || | |||  ||| |
Db    596 ctgtccctctgctgcatgccc 616


RESULT  25
US-08-984-709A-49
; Sequence 49, Application US/08984709A
; Patent No. 6320032
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Mark E.
;    APPLICANT:  Stauderman, Kenneth A.
;    APPLICANT:  Harpold, Michael M.
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Heller Ehrman White & McAuliffe
;      STREET:  4250 Executive Square, Suite 700
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/984,709A
;      FILING DATE:  02-DEC-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  24735-9815 (formerly 6362-9815)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 450-8400
;      TELEFAX:  (619) 587-5360
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7898 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  249...7307
;      OTHER INFORMATION:
US-08-984-709A-49


  Query Match             3.7%;  Score 33;  DB 4;  Length 7898;
  Best Local Similarity   51.7%;  Pred. No. 6.8;
  Matches   75;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    481 atcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcccaccggcatacaaggg 540
          | |  || ||||| |   |   | | | | | | | || ||  | | | |    | | |
Db   6489 AGCCGGCCGGCGGCCCCAGGCGGAGAGGAGGCCGAGGCCTCGGACCCAGCCGACGAGGAG 6548

Qy    541 gtcaaacaaatcatcaaggccctgcgcgcgggcgaggcaaccatcatcctgcccgaccac 600
          ||||  || |||| ||   ||    || | | || | ||   |   | | | | || |
Db   6549 GTCAGCCACATCACCAGCTCCGCCTGCCCCTGGCAGCCCACAGCCGAGCCCCATGGCCCC 6608

Qy    601 gtcccttctccgcaggaaggcggcg 625
          |   | ||||||  ||  |||||||
Db   6609 GAAGCCTCTCCGGTGGCCGGCGGCG 6633


RESULT  26
US-08-541-033A-25
; Sequence 25, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/541,033A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Whitlock, Ted W.
;       REGISTRATION NUMBER:  36,965
;       REFERENCE/DOCKET NUMBER:  UF155
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (904) 375-8100
;       TELEFAX:  (904) 372-5800
;   INFORMATION FOR SEQ ID NO:  25:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1473 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  CDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  4..1431
US-08-541-033A-25


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 1473;
  Best Local Similarity   45.4%;  Pred. No. 3.7;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db     972 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1031

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
           |   || |   |  | | | |  ||| | | ||     |    | |    | ||
Db    1032 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1091

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
           |    ||||  ||      | || |    ||   | |    ||| |||  | |  |
Db    1092 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1151

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   ||||| | ||
Db    1152 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1211

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1212 CGCCGGCGGCGTGGCGGTCA 1231


RESULT  27
US-08-828-451-25
; Sequence 25, Application US/08828451
; Patent No. 5985634
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE - AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/828,451
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/541,033
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Whitlock, Ted W.
;       REGISTRATION NUMBER:  36,965
;       REFERENCE/DOCKET NUMBER:  UF155
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (904) 375-8100
;       TELEFAX:  (904) 372-5800
;   INFORMATION FOR SEQ ID NO:  25:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1473 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  CDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  4..1431
US-08-828-451-25


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 1473;
  Best Local Similarity   45.4%;  Pred. No. 3.7;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db     972 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1031

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
           |   || |   |  | | | |  ||| | | ||     |    | |    | ||
Db    1032 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1091

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
           |    ||||  ||      | || |    ||   | |    ||| |||  | |  |
Db    1092 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1151

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   ||||| | ||
Db    1152 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1211

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1212 CGCCGGCGGCGTGGCGGTCA 1231


RESULT  28
US-08-541-033A-23
; Sequence 23, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE `- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1506 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4..1464
US-08-541-033A-23


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 1506;
  Best Local Similarity   45.4%;  Pred. No. 3.7;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1005 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1064

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | |  | |  ||| | | ||     |    |  |    | ||
Db    1065 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1124

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| |||  | |  |
Db    1125 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1184

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   |||||  | ||
Db    1185 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1244

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1245 CGCCGGCGGCGTGGCGGTCA 1264


RESULT  29
US-08-828-451-23
; Sequence 23, Application US/08828451
; Patent No. 5985634
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE - AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/828,451
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1506 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4..1464
US-08-828-451-23


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 1506;
  Best Local Similarity   45.4%;  Pred. No. 3.7;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1005 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1064

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | |  | |  ||| | | ||     |    |  |    | ||
Db    1065 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1124

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| |||  | |  |
Db    1125 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1184

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   |||||  | ||
Db    1185 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1244

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1245 CGCCGGCGGCGTGGCGGTCA 1264


RESULT  30
US-08-541-033A-7
; Sequence 7, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
```

```
;       ZIP:  32606
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/541,033A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Whitlock, Ted W.
;       REGISTRATION NUMBER:  36,965
;       REFERENCE/DOCKET NUMBER:  UF155
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (904) 375-8100
;       TELEFAX:  (904) 372-5800
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1969 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-541-033A-7


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 1969;
  Best Local Similarity   45.4%;  Pred. No. 4.2;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||      |  | | ||   || || |  | |     | ||| ||  |
Db     981 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1040

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | | | | ||| | | ||      |    |  |    | ||
Db    1041 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1100

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| |||  | |  |
Db    1101 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1160

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| |||   |   | ||   |||||  | ||
Db    1161 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1220

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1221 CGCCGGCGGCGTGGCGGTCA 1240


RESULT  31
US-08-828-451-7
; Sequence 7, Application US/08828451
; Patent No. 5985634
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE - AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/828,451
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/541,033
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Whitlock, Ted W.
;       REGISTRATION NUMBER:  36,965
;       REFERENCE/DOCKET NUMBER:  UF155
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (904) 375-8100
;       TELEFAX:  (904) 372-5800
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1969 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-828-451-7


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 1969;
  Best Local Similarity   45.4%;  Pred. No. 4.2;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||      |  | | ||   || || |  | |     | ||| ||  |
Db     981 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1040

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | | | | ||| | | ||      |    |  |    | ||
Db    1041 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1100

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| |||  | |  |
Db    1101 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1160

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| |||   |   | ||   |||||  | ||
Db    1161 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1220

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1221 CGCCGGCGGCGTGGCGGTCA 1240


RESULT  32
US-08-541-033A-19
; Sequence 19, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/541,033A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Whitlock, Ted W.
;       REGISTRATION NUMBER:  36,965
;       REFERENCE/DOCKET NUMBER:  UF155
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (904) 375-8100
;       TELEFAX:  (904) 372-5800
;   INFORMATION FOR SEQ ID NO:  19:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2096 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  CDNA
US-08-541-033A-19


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 2096;
  Best Local Similarity   45.4%;  Pred. No. 4.3;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1106 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1165

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | |  | |  ||| | | ||     |   |  |    | ||
Db    1166 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1225

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| ||| | |  |
Db    1226 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1285

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   |||||  | ||
Db    1286 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1345

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1346 CGCCGGCGGCGTGGCGGTCA 1365


RESULT  33
US-08-828-451-19
; Sequence 19, Application US/08828451
; Patent No. 5985634
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE - AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/828,451
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2096 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
US-08-828-451-19


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 2096;
  Best Local Similarity   45.4%;  Pred. No. 4.3;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1106 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1165

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | |  | |  ||| | | ||     |   |  |    | ||
Db    1166 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1225

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| ||| | |  |
Db    1226 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1285

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   |||||  | ||
Db    1286 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1345

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1346 CGCCGGCGGCGTGGCGGTCA 1365


RESULT  34
US-08-541-033A-3
; Sequence 3, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE `- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2099 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1568
US-08-541-033A-3


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 2099;
  Best Local Similarity   45.4%;  Pred. No. 4.3;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1109 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1168

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
           |   || |   |   | |  | |  ||| | | ||     |    |  |    | ||
Db    1169 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1228

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
           |     ||||  ||       | || |    ||   | |    ||| ||| | | |
Db    1229 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1288

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||  |||||  | ||
Db    1289 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1348

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1349 CGCCGGCGGCGTGGCGGTCA 1368


RESULT  35
US-08-828-451-3
; Sequence 3, Application US/08828451
; Patent No. 5985634
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE - AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/828,451
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2099 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1568
US-08-828-451-3


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 2099;
  Best Local Similarity   45.4%;  Pred. No. 4.3;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1109 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1168

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
           |   || |   |   | |  | |  ||| | | ||     |    |  |    | ||
Db    1169 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1228

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
           |     ||||  ||       | || |    ||   | |    ||| ||| | | |
Db    1229 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1288

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||  |||||  | ||
Db    1289 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1348

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1349 CGCCGGCGGCGTGGCGGTCA 1368


RESULT  36
US-08-541-033A-18
; Sequence 18, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
US-08-541-033A-18


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 2137;
  Best Local Similarity   45.4%;  Pred. No. 4.4;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1148 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1207

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | |  | |  ||| | | ||     |    | |    | ||
Db    1208 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1267

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| |||  | |  |
Db    1268 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1327

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   ||||| | ||
Db    1328 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1387

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1388 CGCCGGCGGCGTGGCGGTCA 1407


RESULT  37
US-08-828-451-18
; Sequence 18, Application US/08828451
; Patent No. 5985634
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE - AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/828,451
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
US-08-828-451-18


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 2137;
  Best Local Similarity   45.4%;  Pred. No. 4.4;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1148 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1207

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | |  | |  ||| | | ||     |    | |    | ||
Db    1208 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1267

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| |||  | |  |
Db    1268 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1327

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   ||||| | ||
Db    1328 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1387

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1388 CGCCGGCGGCGTGGCGGTCA 1407


RESULT  38
US-08-541-033A-1
; Sequence 1, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1610
US-08-541-033A-1


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 2140;
  Best Local Similarity   45.4%;  Pred. No. 4.4;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1151 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1210

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | |  | |  ||| | | ||     |   |   | |    | ||
Db    1211 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1270

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| ||| | |  |
Db    1271 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1330

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   |||||  | ||
Db    1331 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1390

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1391 CGCCGGCGGCGTGGCGGTCA 1410


RESULT  39
US-08-828-451-1
; Sequence 1, Application US/08828451
; Patent No. 5985634
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE - AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/828,451
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1610
US-08-828-451-1


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 2140;
  Best Local Similarity   45.4%;  Pred. No. 4.4;
  Matches  118;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     285 ggaagacatcgaaacaatgttcaaagcggtacacggctgggaacacgtgcagcaggcttt 344
           | |  | | ||| ||     |  | | ||   || || |  | |    | ||| ||  |
Db    1151 GTACAAGAGCGACACCGCCGTGTATGTGGGCGACCGCCGCAAGCCTTGGGAGCTGGACTG 1210

Qy     345 ggacaagggcgaagggctgctgttcatcacgccgcacatcggcagctacgatttgggcgg 404
             |   || |   |   | |  | |  ||| | | ||     |   |   | |    | ||
Db    1211 CCAGGTGGACATCGCCTTCCCCTGCGCCACCCAGAACGAGATCGATGAGCACGACGCCGA 1270

Qy     405 acgctacatcagccagcagcttccgttccacctgaccgccatgtacaagccgccgaaaat 464
             |    ||||  ||       | || |    ||   | |    ||| ||| | |  |
Db    1271 GCTGCTGATCAAGCACGGCTGCCAGTACGTGGTGGAGGGCGCCAACATGCCCTCCACCAA 1330

Qy     465 caaagcgatagacaaaatcatgcaggcgggcagggtgcgcggcaaaggcaaaaccgcgcc 524
           | | || ||  ||||   ||   |||| ||||   |   | ||   |||||  | ||
Db    1331 CGAGGCCATCCACAAGTACAACAAGGCCGGCATCATCTACTGCCCCGGCAAGGCGGCCAA 1390

Qy     525 caccggcatacaaggggtca 544
           | |||||      | |||||
Db    1391 CGCCGGCGGCGTGGCGGTCA 1410


RESULT  40
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1


  Query Match             3.7%;  Score 32.8;  DB 4;  Length 15872;
  Best Local Similarity   56.5%;  Pred. No. 11;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     520 gcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcgggcgaggca 579
           ||  || ||||  |    || ||||   | ||  || |||| |||||    |||  |
Db    9548 gccgccgccggtgtcggcggcgtcatcaagatgctccaggcgctgcggcacggcttgctg 9607
```

```
Qy    580 accatcatcctgcccgaccacgtcccttctccgcaggaaggcggcggc 627
          ||  || ||| | || | |||  ||  | ||||| |  | | |  ||
Db   9608 ccccgcaccctccacgccgacgagccgaccccgcatgccgactggagc 9655


RESULT  41
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1


  Query Match             3.7%;  Score 32.8;  DB 2;  Length 43280;
  Best Local Similarity   56.5%;  Pred. No. 17;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    520 gcgcccaccggcatacaaggggtcaaacaaatcatcaaggccctgcgcgcgggcgaggca 579
          ||  || ||||  |    || ||||   | ||  || |||| |||||    |||  |
Db  37410 GCCGCCGCCGGTGTCGGCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTG 37469

Qy    580 accatcatcctgcccgaccacgtcccttctccgcaggaaggcggcggc 627
          ||  || ||| | || | |||  ||  | ||||| |  | | |  ||
```

```
Db  37470 CCCCGCACCCTCCACGCCGACGAGCCGACCCCGCATGCCGACTGGAGC 37517


RESULT  42
US-08-997-080-179
; Sequence 179, Application US/08997080
; Patent No. 5968524
;  GENERAL INFORMATION:
;    APPLICANT:  WATSON, JAMES D.
;    APPLICANT:  TAN, PAUL L.J.
;    TITLE OF INVENTION:  METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
;    NUMBER OF SEQUENCES:  194
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,080
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sleath, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000.1007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  179:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  520 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-997-080-179


  Query Match             3.6%;  Score 32.6;  DB 2;  Length 520;
  Best Local Similarity   49.7%;  Pred. No. 2.7;
  Matches   83;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy    303 gttcaaagcggtacacggctgggaacacgtgcagcaggctttggacaagggcgaagggct 362
          | |||  | ||  | | || | || |  ||  |||||||   || |||||||||| |
Db    312 GCTCACCGAGGAGCGCCGCCGCGACCTGGTCAAGCAGGCCAAGGCCAAGGGCGAGGACGC 371

Qy    363 gctgttcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagca 422
          | |      ||   ||||| |  || ||||| ||  |        ||||   | |
Db    372 CAAGGTGTCGGTGCGCAACATCCGTCGCAACGATATGAACACCTTTCGCATCGCACCGGT 431

Qy    423 gcttccgttccacctgaccgccatgtacaagccgccgaaaatcaaag 469
          |  | |      |   |  |   ||| |||| | ||  |||  ||
Db    432 ACGGCTGCCGACGCCACCGCCGTCGTAGAAGCGACAGAGGATCGCAG 478


RESULT  43
US-08-997-362-179
; Sequence 179, Application US/08997362
; Patent No. 5985287
;  GENERAL INFORMATION:
;    APPLICANT:  Tan, Paul
;    APPLICANT:  Hiyama, Jun
;    APPLICANT:  Visser, Elizabeth
```

```
;    APPLICANT:  Skinner, Margot
;    APPLICANT:  Scott, Linda
;    APPLICANT:  Prestidge, Ross
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
;    NUMBER OF SEQUENCES:  194
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,362
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. Patent Application No. 5985287 08/873,970
;      FILING DATE:  June 12, 1997
;      APPLICATION NUMBER:  U.S. Patent Application No. 5985287 08/705,347
;      FILING DATE:  August 29, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sleath, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000.1002c2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  179:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  520 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-997-362-179


  Query Match             3.6%;  Score 32.6;  DB 2;  Length 520;
  Best Local Similarity   49.7%;  Pred. No. 2.7;
  Matches   83;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy    303 gttcaaagcggtacacggctgggaacacgtgcagcaggctttggacaagggcgaagggct 362
          | |||  | ||  | | || | || |  ||  |||||||   || |||||||| |
Db    312 GCTCACCGAGGAGCGCCGCCGCGACCTGGTCAAGCAGGCCAAGGCCAAGGGCGAGGACGC 371

Qy    363 gctgttcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagca 422
            | |      ||   ||||| |  || ||||| ||  |        ||||   | |
Db    372 CAAGGTGTCGGTGCGCAACATCCGTCGCAACGATATGAACACCTTTCGCATCGCACCGGT 431

Qy    423 gcttccgttccacctgaccgccatgtacaagccgccgaaaatcaaag 469
            |  | ||      |  | |   ||| ||||  | ||  |||  ||
Db    432 ACGGCTGCCGACGCCACCGCCGTCGTAGAAGCGACAGAGGATCGCAG 478


RESULT  44
US-09-095-855-179
; Sequence 179, Application US/09095855
; Patent No. 6160093
;  GENERAL INFORMATION:
;    APPLICANT:  Tan, Paul
;    APPLICANT:  Visser, Elizabeth
;    APPLICANT:  Skinner, Margot
;    APPLICANT:  Prestidge, Ross
;    TITLE OF INVENTION:  Compounds and Methods for
;    TITLE OF INVENTION:  Treatment and Diagnosis of Mycobacterial Infections
;    NUMBER OF SEQUENCES:  208
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/095,855
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/705,347
;      FILING DATE:  29-AUG-1996
;      APPLICATION NUMBER:  08/873,970
;      FILING DATE:  12-JUN-1997
;      APPLICATION NUMBER:  08/997,362
;      FILING DATE:  23-DEC-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sleath, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000.1002c3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  179:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  520 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-095-855-179


  Query Match             3.6%;  Score 32.6;  DB 4;  Length 520;
  Best Local Similarity   49.7%;  Pred. No. 2.7;
  Matches   83;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy    303 gttcaaagcggtacacggctgggaacacgtgcagcaggctttggacaagggcgaagggct 362
          | |||  | ||  | | || | || |  ||  |||||||   || |||||||| |
Db    312 GCTCACCGAGGAGCGCCGCCGCGACCTGGTCAAGCAGGCCAAGGCCAAGGGCGAGGACGC 371

Qy    363 gctgttcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagca 422
            | |      ||   ||||| |  || ||||| ||  |        ||||   | |
Db    372 CAAGGTGTCGGTGCGCAACATCCGTCGCAACGATATGAACACCTTTCGCATCGCACCGGT 431

Qy    423 gcttccgttccacctgaccgccatgtacaagccgccgaaaatcaaag 469
            |  | ||      |  | |   ||| ||||  | ||  |||  ||
Db    432 ACGGCTGCCGACGCCACCGCCGTCGTAGAAGCGACAGAGGATCGCAG 478


RESULT  45
US-09-324-542-179
; Sequence 179, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Tan, Paul L.J.
;  APPLICANT: Prestidge, Ross
;  TITLE OF INVENTION: Methods and Compounds for the Treatment
;  TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
;  FILE REFERENCE: 11000.1007c1
;  CURRENT APPLICATION NUMBER: US/09/324,542
;  CURRENT FILING DATE: 1999-06-02
;  EARLIER APPLICATION NUMBER: US 08/997,080
;  EARLIER FILING DATE: 1997-12-23
;  NUMBER OF SEQ ID NOS: 194
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 179
;   LENGTH: 520
;   TYPE: DNA
;   ORGANISM: Mycobacterium vaccae
US-09-324-542-179


  Query Match             3.6%;  Score 32.6;  DB 4;  Length 520;
  Best Local Similarity   49.7%;  Pred. No. 2.7;
  Matches   83;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     303 gttcaaagcggtacacggctgggaacacgtgcagcaggctttggacaagggcgaagggct 362
           | |||  | ||  | | || | || |  ||  |||||||   || ||||||||| |
Db     312 gctcaccgaggagcgccgccgcgacctggtcaagcaggccaaggccaagggcgaggacgc 371

Qy     363 gctgttcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagca 422
             | |      ||   ||||| |  || ||||| ||  |        ||||   | |
Db     372 caaggtgtcggtgcgcaacatccgtcgcaacgatatgaacacctttcgcatcgcaccggt 431

Qy     423 gcttccgttccacctgaccgccatgtacaagccgccgaaaatcaaag 469
            |  | |      |   |  |   ||| ||||  | ||  |||  ||
Db     432 acggctgccgacgccaccgccgtcgtagaagcgacagaggatcgcag 478


Search completed: July  1, 2002, 04:28:32
Job time: 113437 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2002, 01:19:33 ; Search time 415.2 Seconds
(without alignments)
530.668 Million cell updates/sec

Title: US-09-303-518D-569
Perfect score: 897
Sequence: 1 atgtttcgtttacaattcag.....accgctacaaaatgccgtaa 897

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	36	4.0	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	35.6	4.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	3	33.8	3.8	1838	3	US-08-470-260-7	Sequence 7, Appli
	4	33.8	3.8	1838	3	US-08-471-491-7	Sequence 7, Appli
	5	33.8	3.8	1838	3	US-08-466-662-7	Sequence 7, Appli
	6	32.2	3.6	289	4	US-09-007-005-17	Sequence 17, Appl
	7	32.2	3.6	289	4	US-09-244-796-17	Sequence 17, Appl
	8	32.2	3.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	9	32.2	3.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	10	32	3.6	1619	2	US-08-991-946A-2	Sequence 2, Appli
	11	31.8	3.5	1483	4	US-09-262-749-1	Sequence 1, Appli
	12	31.8	3.5	1553	3	US-09-022-669-1	Sequence 1, Appli
c	13	31.6	3.5	289	4	US-09-007-005-17	Sequence 17, Appl
c	14	31.6	3.5	289	4	US-09-244-796-17	Sequence 17, Appl
	15	31.6	3.5	496	4	US-08-905-223-23	Sequence 23, Appl
	16	31.6	3.5	496	4	US-09-247-155-23	Sequence 23, Appl
	17	31.4	3.5	1119	4	US-08-987-943-2	Sequence 2, Appli
	18	31.4	3.5	1755	4	US-08-987-943-1	Sequence 1, Appli
c	19	31.4	3.5	5247	1	US-08-920-812-15	Sequence 15, Appl
c	20	31.4	3.5	5247	1	US-08-920-827-15	Sequence 15, Appl
c	21	31.4	3.5	5247	1	US-08-921-177-15	Sequence 15, Appl
c	22	31.4	3.5	5247	1	US-08-362-577C-15	Sequence 15, Appl
c	23	31.4	3.5	5247	2	US-08-920-828-15	Sequence 15, Appl
c	24	31.2	3.5	11958	4	US-09-134-246-8	Sequence 8, Appli
	25	31	3.5	480	3	US-08-814-052-15	Sequence 15, Appl
c	26	31	3.5	1082	1	US-08-715-252-1	Sequence 1, Appli
c	27	31	3.5	1082	2	US-08-453-051-3	Sequence 3, Appli
c	28	31	3.5	1130	1	US-08-406-948A-5	Sequence 5, Appli
	29	30	3.3	3358	3	US-09-248-571-2	Sequence 2, Appli
	30	30	3.3	6816	4	US-09-404-650-1	Sequence 1, Appli
	31	30	3.3	6855	4	US-09-404-650-3	Sequence 3, Appli
c	32	29.8	3.3	2478	1	US-08-481-626-1	Sequence 1, Appli
c	33	29.8	3.3	4020	4	US-09-050-159-130	Sequence 130, App
c	34	29.8	3.3	4024	4	US-09-162-484-18	Sequence 18, Appl
c	35	29.6	3.3	2256	1	US-07-794-393-1	Sequence 1, Appli
c	36	29.6	3.3	2256	1	US-08-001-711-1	Sequence 1, Appli
	37	29.4	3.3	11219	1	US-07-642-734C-1	Sequence 1, Appli
	38	29.4	3.3	11219	3	US-08-439-009A-1	Sequence 1, Appli
	39	29.2	3.3	342	4	US-08-905-223-196	Sequence 196, App
	40	29.2	3.3	1432	1	US-07-914-282D-3	Sequence 3, Appli
	41	29.2	3.3	1432	1	US-08-276-887A-3	Sequence 3, Appli
	42	29.2	3.3	1432	5	PCT-US93-02460-3	Sequence 3, Appli
	43	29.2	3.3	1683	1	US-07-945-283-3	Sequence 3, Appli
c	44	29.2	3.3	1716	4	US-09-321-981-4	Sequence 4, Appli
	45	29.2	3.3	1782	2	US-08-714-168-3	Sequence 3, Appli
	46	29.2	3.3	1782	4	US-09-320-721A-3	Sequence 3, Appli
c	47	29.2	3.3	1831	6	5215881-1	Patent No. 5215881
c	48	29.2	3.3	1831	6	5215881-3	Patent No. 5215881
c	49	29.2	3.3	2223	1	US-08-317-522A-8	Sequence 8, Appli
c	50	29.2	3.3	2223	1	US-08-439-818A-8	Sequence 8, Appli
c	51	29.2	3.3	2223	2	US-08-751-965-8	Sequence 8, Appli
c	52	29.2	3.3	2223	2	US-08-738-975-8	Sequence 8, Appli
c	53	29.2	3.3	2223	2	US-08-728-626-8	Sequence 8, Appli
c	54	29.2	3.3	2223	3	US-08-808-599A-8	Sequence 8, Appli
	55	29	3.2	1095	1	US-08-385-186-8	Sequence 8, Appli
	56	29	3.2	1425	1	US-08-385-186-12	Sequence 12, Appl
	57	29	3.2	1639	1	US-08-334-698-5	Sequence 5, Appli
	58	29	3.2	1639	1	US-08-228-932-5	Sequence 5, Appli
	59	29	3.2	1639	1	US-08-468-939-5	Sequence 5, Appli
	60	29	3.2	1639	2	US-08-406-855A-5	Sequence 5, Appli
	61	29	3.2	1639	2	US-08-722-190-5	Sequence 5, Appli
	62	29	3.2	1639	3	US-08-244-354-5	Sequence 5, Appli
	63	29	3.2	1639	3	US-09-206-899-5	Sequence 5, Appli
	64	29	3.2	1639	5	PCT-US95-04203-5	Sequence 5, Appli
c	65	29	3.2	7122	4	US-09-318-448-2	Sequence 2, Appli
	66	29	3.2	11492	4	US-08-991-840A-1	Sequence 1, Appli
	67	29	3.2	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	68	28.8	3.2	4800	3	US-08-941-445A-4	Sequence 4, Appli
	69	28.8	3.2	6226	1	US-08-542-363-1	Sequence 1, Appli
	70	28.8	3.2	6226	4	US-09-100-089-1	Sequence 1, Appli
	71	28.8	3.2	35081	2	US-08-752-760A-1	Sequence 1, Appli
	72	28.8	3.2	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	73	28.6	3.2	1424	1	US-08-403-634-3	Sequence 3, Appli
	74	28.6	3.2	1424	4	US-08-913-441B-3	Sequence 3, Appli
	75	28.6	3.2	1611	5	PCT-US93-08743-3	Sequence 3, Appli
c	76	28.6	3.2	5925	3	US-08-470-260-4	Sequence 4, Appli
c	77	28.6	3.2	5925	3	US-08-471-491-4	Sequence 4, Appli
c	78	28.6	3.2	5925	3	US-08-466-662-4	Sequence 4, Appli
c	79	28.6	3.2	10299	2	US-08-477-451-1	Sequence 1, Appli
	80	28.6	3.2	10299	2	US-08-477-451-5	Sequence 5, Appli
c	81	28.6	3.2	19932	2	US-08-477-451-25	Sequence 25, Appl
c	82	28.4	3.2	434	4	US-09-060-756-67	Sequence 67, Appl
	83	28.4	3.2	476	3	US-08-050-259B-19	Sequence 19, Appl
	84	28.4	3.2	1278	4	US-09-613-182-14	Sequence 14, Appl
	85	28.4	3.2	2846	4	US-09-613-182-5	Sequence 5, Appli
c	86	28.4	3.2	2963	4	US-09-232-200-60	Sequence 60, Appl
c	87	28.4	3.2	2963	4	US-09-232-197-60	Sequence 60, Appl
c	88	28.4	3.2	2963	4	US-09-232-201-60	Sequence 60, Appl
c	89	28.4	3.2	4529	1	US-08-565-386-1	Sequence 1, Appli
c	90	28.2	3.1	3416	2	US-08-701-240-3	Sequence 3, Appli
c	91	28.2	3.1	3416	3	US-09-138-236-3	Sequence 3, Appli
c	92	28.2	3.1	14683	2	US-08-819-866-1	Sequence 1, Appli
c	93	28.2	3.1	14683	2	US-09-023-715-1	Sequence 1, Appli
	94	28	3.1	1353	4	US-09-011-197-3	Sequence 3, Appli
	95	28	3.1	1539	3	US-08-582-740-69	Sequence 69, Appl
	96	28	3.1	1539	4	US-09-109-879-69	Sequence 69, Appl
	97	28	3.1	1624	3	US-08-582-740-67	Sequence 67, Appl
	98	28	3.1	1624	4	US-09-109-879-67	Sequence 67, Appl
	99	28	3.1	2361	4	US-09-011-197-1	Sequence 1, Appli
	100	28	3.1	3636	3	US-09-074-579-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT   1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             4.0%;  Score 36;  DB 1;  Length 7218;
  Best Local Similarity   3.2%;  Pred. No. 0.29;
  Matches   12;  Conservative  199;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     211 aaaacggtcaaagccgtttttgcggaaacggcaaaaggcggtttggaacttgccccccgcg 270
           |||  | |  |||   ||  | |::::: :: :::::: ::   ::::   :     : :
Db    1457 AAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1398

Qy     271 tttttcagaaaaccggaagacatagaaacaatgttcaaagcggtacacggctgggaacat 330
             ::::::  :::::: : ::::: :: :   :::: :: : : ::  ::::: :
Db    1397 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1338

Qy     331 gtgcagcaggctttggacaaacacgaagggctgctattcatcacgccgcacatcggcagc 390
           : : :: :::    ::: ::: : ::::::  :  :   :  : : : : : :  :: ::
Db    1337 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1278

Qy     391 tacgatttgggcggacgctacatcagccagcagcttccgttcccgctgaccgccatgtac 450
            : ::   ::: ::: :  : :  ::  :: ::     :     :  ::  :  : : :
```

```
Db    1277 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1218

Qy     451 aaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggttcgcggcaaa 510
           :::  :  :::::  :::: :: ::: ::::  : : ::: ::: ::::   : :: :::
Db    1217 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1158

Qy     511 ggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcg 570
           :::::::  : :   :  :: : : ::::::  ::: :::  :  ::::   : :   :
Db    1157 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1098

Qy     571 ggcgaagcaa 580
           :: :::: ::
Db    1097 RRRRRRRRRR 1088


RESULT   2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


  Query Match             4.0%;  Score 35.6;  DB 4;  Length 4411529;
  Best Local Similarity   50.0%;  Pred. No. 12;
  Matches   89;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy     651 caaacctgcctataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaac 710
           |||  |   |   | | || || |  | ||  ||  |||||  | | | |         |
Db  498793 CAAGACCATCACCAACCTGGCGATCACCGCCCAACAGGCACGGGCCGACGAGACACTGTC 498734

Qy     711 cctgtttttctgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccc 770
            ||| |     || ||||     |    ||   || |  ||| ||  || |||| | |
Db  498733 GCTGATCCGGCGCGGCGACGAAGAGGTCCGCAAGCAGGCCTTCTATCAGCGCATCGACGC 498674

Qy     771 cgtccaaggggaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaat 828
           | | ||| |  |  | |||| | |||  ||||       ||||||   ||| | | ||
Db  498673 CATGCAACGCCAGCTCAACGACTACATGGCCCGCCGCCACGCCGTCGACAAACCCGAT 498616


RESULT   3
US-08-470-260-7
; Sequence 7, Application US/08470260
; Patent No. 6077706
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    APPLICANT:  Bugnoli, Massimo
;    APPLICANT:  Telford, John
;    APPLICANT:  Macchia, Giovanni
;    APPLICANT:  Rappuoli, Rino
;    TITLE OF INVENTION:  Helicobacter Pylori Proteins Useful
;    TITLE OF INVENTION:  for Vaccines and Diagnostics
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
```

```
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94608-2916
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/470,260
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/256,848
;       FILING DATE:  21-OCT-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McClung, Barbara G.
;       REGISTRATION NUMBER:  33,113
;       REFERENCE/DOCKET NUMBER:  0316.001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 601-2708
;       TELEFAX:  (510) 655-3542
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1838 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-470-260-7


  Query Match             3.8%;  Score 33.8;  DB 3;  Length 1838;
  Best Local Similarity   45.1%;  Pred. No. 0.7;
  Matches  125;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy     546 acaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtccc 605
           | ||| |    |||    |  ||  ||| ||||    |   |  ||   ||| | | |
Db     981 AGAAAACGCTGAAGTGGAGTTTTTAGGCAAAGCTGGAAGGATTGTGATTGACAAAGACAA 1040

Qy     606 ctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcctatac 665
           | || |    | ||  ||  |||||    | |  ||| |    | || |   ||  | |
Db    1041 CACCACGATCGTAGATGGCAAAGGCCATAGCGATGATGTTAAAGACAGAGTCGCGCAGAT 1100

Qy     666 catgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctg 725
           ||  || |    ||||    | ||  ||   | | |   |||   ||       | |
Db    1101 CAAAACCCAAATTGCAAGTACGACAAGCGATTATGACAAAGAAAAATTGCAAGAAAGATT 1160

Qy     726 cgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggggaatt 785
           |     ||  |||||||    | | |  |       |    |     ||| |||| |
Db    1161 GGCTAAACTCTCTGGCGGTGTGGCTGTGATTAAAGTGGGCGCTGCGAGTGAAGTGGAAAT 1220

Qy     786 gaacggcgacaaagcccatgatgccgccgtgttcaac 822
           ||| |   | |||| ||  |  |  | || ||| | |
Db    1221 GAAAGAGAAAAAAGACCGGGTGGATGACGCGTTGAGC 1257


RESULT   4
US-08-471-491-7
; Sequence 7, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0044
;  CURRENT APPLICATION NUMBER: US/08/471,491B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 1838
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
US-08-471-491-7


  Query Match             3.8%;  Score 33.8;  DB 3;  Length 1838;
  Best Local Similarity   45.1%;  Pred. No. 0.7;
  Matches  125;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy     546 acaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtccc 605
           | ||| |    |||    |  ||  ||| ||||    |   |  ||   ||| | | |
Db     981 agaaaacgctgaagtggagtttttaggcaaagctggaaggattgtgattgacaaagacaa 1040

Qy     606 ctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcctatac 665
           | || |    | ||  ||  |||||    | |  ||| |    | || |   ||  | |
Db    1041 caccacgatcgtagatggcaaaggccatagcgatgatgttaaagacagagtcgcgcagat 1100

Qy     666 catgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctg 725
           ||  || |    ||||    | ||  ||   | | |   |||   ||       | |
Db    1101 caaaacccaaattgcaagtacgacaagcgattatgacaaagaaaaattgcaagaaagatt 1160

Qy     726 cgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggggaatt 785
           |     ||  |||||||    | | |  |       |    |     ||| |||| |
Db    1161 ggctaaactctctggcggtgtggctgtgattaaagtgggcgctgcgagtgaagtggaaat 1220

Qy     786 gaacggcgacaaagcccatgatgccgccgtgttcaac 822
           ||| |   | |||| ||  |  |  | || ||| | |
Db    1221 gaaagagaaaaaagaccgggtggatgacgcgttgagc 1257


RESULT   5
US-08-466-662-7
; Sequence 7, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0057
;  CURRENT APPLICATION NUMBER: US/08/466,662B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 1838
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
US-08-466-662-7


  Query Match             3.8%;  Score 33.8;  DB 3;  Length 1838;
  Best Local Similarity   45.1%;  Pred. No. 0.7;
  Matches  125;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy     546 acaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtccc 605
           | ||| |    |||    |  ||  ||| ||||    |   |  ||   ||| | | |
Db     981 agaaaacgctgaagtggagtttttaggcaaagctggaaggattgtgattgacaaagacaa 1040

Qy     606 ctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcctatac 665
           | || |    | ||  ||  |||||    | |  ||| |    | || |   ||  | |
Db    1041 caccacgatcgtagatggcaaaggccatagcgatgatgttaaagacagagtcgcgcagat 1100

Qy     666 catgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctg 725
           ||  || |    ||||    | ||  ||   | | |   |||   ||       | |
Db    1101 caaaacccaaattgcaagtacgacaagcgattatgacaaagaaaaattgcaagaaagatt 1160
```

```
Qy    726 cgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggggaatt 785
          |     ||  |||||||    | | |   |        |     |     ||| |||| |
Db   1161 ggctaaactctctggcggtgtggctgtgattaaagtgggcgctgcgagtgaagtggaaat 1220

Qy    786 gaacggcgacaaagcccatgatgccgccgtgttcaac 822
          ||| |   | |||| ||  |  |  | || ||| | |
Db   1221 gaaagagaaaaaagaccgggtggatgacgcgttgagc 1257


RESULT   6
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17


  Query Match             3.6%;  Score 32.2;  DB 4;  Length 289;
  Best Local Similarity   9.1%;  Pred. No. 0.82;
  Matches   24;  Conservative   99;  Mismatches  142;  Indels    0;  Gaps    0;

Qy    294 agaaacaatgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaaca 353
          :|:|:|:|    :  :|:|:  |    :   : :| |     |:|: :  : |:  :
Db      5 rgrarcrararururarcrurarururururarcrararururarcrararurgrnrnrsrn 64

Qy    354 cgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacat 413
          : : :  :  :  ::    :  : :::   :::  :::    ::  :     :::
Db     65 rnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrn 124

Qy    414 cagccagcagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgat 473
          ::  : ::     :     :  :  ::    :    :  :: :  :* :  ::
Db    125 rnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrn 184

Qy    474 agacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcat 533
          : ::: :  :: : ::: : ::: :      ::: : : : :   :|:  ::  |:|::
Db    185 rnrsrnrnrsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgrurararcru 244

Qy    534 acaaggggtcaaacaaatcatcaaa 558
          :|: : :| |: :|: |  |  |||
Db    245 rcrururgrcrgrcruaaaaaaaaa 269


RESULT   7
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17


  Query Match             3.6%;  Score 32.2;  DB 4;  Length 289;
  Best Local Similarity   9.1%;  Pred. No. 0.82;
  Matches   24;  Conservative   99;  Mismatches  142;  Indels    0;  Gaps    0;

Qy    294 agaaacaatgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaaca 353
          :|:|:|:|    :  :|:|:  |    :   : :| |     |:|: :  : |:  :
Db      5 rgrarcrararururarcrurarururururarcrararururarcrararurgrnrnrsrn 64

Qy    354 cgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacat 413
          : : :  :  :  ::    :  : :::   :::  :::    ::  :     :::
Db     65 rnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrn 124

Qy    414 cagccagcagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgat 473
          ::  : ::     :     :  :  ::    :    :  :: :  :: :  ::
Db    125 rnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrn 184

Qy    474 agacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcat 533
          : ::: :  :: : ::: : ::: :      ::: : : : :   :|:  ::  |:|::
Db    185 rnrsrnrnrsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgrurararcru 244

Qy    534 acaaggggtcaaacaaatcatcaaa 558
          :|: : :| |: :|: |  |  |||
Db    245 rcrururgrcrgrcruaaaaaaaaa 269


RESULT   8
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/232,463
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/935,313
;       FILING DATE:
;       APPLICATION NUMBER:  EP 91 114 300.6
;       FILING DATE:  26-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)836-9300
;       TELEFAX:  (703)683-4109
;       TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7218 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             3.6%;  Score 32.2;  DB 1;  Length 7218;
  Best Local Similarity   12.4%;  Pred. No. 4.8;
  Matches   31;  Conservative  111;  Mismatches  109;  Indels    0;  Gaps    0;

Qy       4 tttcgtttacaattcaggctgtttccccctttgcgaaccgccatgcacatcctgttgacc 63
           :::: ::: :  :::   :: :::::::::::  :   :: :: : : : :::: ::  ::
Db    1254 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1313

Qy      64 gccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaaccgg 123
           ::::: :::   : :::::::::: :: :: :::::::: ::: : : : ::      ::
Db    1314 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1373

Qy     124 ctcggacatctggcgttttaccttttaaaggaagaccgcgcgcgcatcgtcgccaatatg 183
           :::   : :::  : :::: :::::::          :: : : : : :: :: :: : :
Db    1374 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1433

Qy     184 cgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgtttttgcggaaacggca 243
           :||     |  | |  ||| |     | |   |  | |   ||  || | |  | | |
Db    1434 YGTACCAAATTCTTCTATCTCTTTAACTACTTGCATAGATAGGTAATTACAGTGATGCCT 1493

Qy     244 aaaggcggttt 254
           | | || ||||
Db    1494 ACATGCCGTTT 1504


RESULT   9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             3.6%;  Score 32.2;  DB 4;  Length 4403765;
  Best Local Similarity   50.0%;  Pred. No. 51;
  Matches   89;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy     651 caaacctgcctataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaac 710
           |||  |   |   | | || || |  | ||  ||  |||||  | | | |         |
Db  500235 CAAGACCATCACCAACCTGGCGATCACCGCCCAACAGGCACGGGCCGACGAGACACTGTC 500176

Qy     711 cctgtttttctgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccc 770
           ||| |     || ||||     |    ||   || |  ||| ||  || ||||  | |
Db  500175 GCTGATCCGGCGCGGCGACGAAGAGGTCCGCAAGCAGGCCTTCTATCAGCGCATCGACGC 500116

Qy     771 cgtccaaggggaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaat 828
           | | ||| |  |  | |||| | |||  ||||      ||||||   ||| | | ||
Db  500115 CATGCAACGCCAGCTCAACGACTACATGGCCCGCCGCCACGCCGTCGACAAACCCGAT 500058


RESULT  10
US-08-991-946A-2/c
; Sequence 2, Application US/08991946A
; Patent No. 5945306
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Tang, Tom Y.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  RAS PROTEINS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/991,946A
;      FILING DATE:  December 16, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0445 US
;    TELECOMMUNICATION INFORMATION:
;     . TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1619 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TLYMNOT01
;      CLONE:  143362
US-08-991-946A-2
```

```
  Query Match             3.6%;  Score 32;  DB 2;  Length 1619;
  Best Local Similarity   50.0%;  Pred. No. 2.4;
  Matches   80;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     433 ccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcg 492
           || || ||  ||    ||   || ||  |      |   | | |||   ||  ||||||
Db    1593 CCACTCACGCCCCCCCACCCCCCCCCCCAGCGTTCATTAACAAACAGCCTCCAGCAGGCC 1534

Qy     493 ggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatc 552
           ||    ||||||  | | |    | | | ||| ||  |    ||      || ||| |
Db    1533 AGCTTTCTTCGCGCGACACGGGGCAGCACCCCTTCCCTCTCCAAAACAACCACACAGAGG 1474

Qy     553 atcaaagccctgcgttcgggcgaagcaaccatcgtcctgc 592
           |  ||   ||  ||  ||||||  ||  |  || |||| |
Db    1473 AGGAACCACCACCGCGCGGGCGGCGCCTCTCTCCTCCTCC 1434


RESULT  11
US-09-262-749-1
; Sequence 1, Application US/09262749
; Patent No. 6261793
;  GENERAL INFORMATION:
;    APPLICANT:  Whyte, David
;    APPLICANT:  McGuirk, Marnie
;    APPLICANT:  Nunez-Oliva, Irma
;    APPLICANT:  Hockenberry, Tish
;    APPLICANT:  Pai, James
;    TITLE OF INVENTION:  RAS CONVERTING ENDOPROTEASE (RCE) AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering-Plough Corporation
;      STREET:  2000 Galloping Hill Road  K-6-1-1990
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Power Macintosh
;      OPERATING SYSTEM:  MACOS
;      SOFTWARE:  MicroSoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/262,749
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thampoe, Immac J
;      REGISTRATION NUMBER:  36,322
;      REFERENCE/DOCKET NUMBER:  OC01005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 298-5061
;      TELEFAX:  (908) 298-5388
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1483 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14..1000
US-09-262-749-1


  Query Match             3.5%;  Score 31.8;  DB 4;  Length 1483;
  Best Local Similarity   49.1%;  Pred. No. 2.7;
  Matches   84;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy      16 ttcaggctgtttccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaa 75
           |||||||||      |  ||| | | | ||  |||    |||||||    || || |
Db     332 TTCAGGCTGGAGGGCATTTTCCTAGCGGCGCTGCTGCCCCTGTTGCTGACCATGATTCTT 391

Qy      76 tgcctctccctgctgccgctttcctgtctgcacacgctgggaaaccggctcggacatctg 135
           | |||   ||  |||  ||    || | || |     |   |  ||| |   |     |||
Db     392 TTCCTGGGCCCACTGATGCAGCTCTCTATGGATTGCCCTTGTGACCTGGCAGATGGGCTG 451

Qy     136 gcgttttacctttaaaggaagaccgcgcgcgcatcgtcgccaatatgcgt 186
           |  ||  |||            | | || |||  | || |  | ||||||
Db     452 AAGGTTGTCCTGGCCCCCCGCTCCTGGGCCCGCTGCCTCACAGACATGCGT 502


RESULT  12
US-09-022-669-1
; Sequence 1, Application US/09022669
; Patent No. 6110717
;  GENERAL INFORMATION:
;    APPLICANT:  KIKLY, KRISTINE
;    APPLICANT:  SOUTHAN, CHRISTOPHER
;    APPLICANT:  KNAB, ANNE
;    TITLE OF INVENTION:  Human RCE1
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/022,669
;      FILING DATE:  12-FEB-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  97304437.3
;      FILING DATE:  JUNE 24, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-70379
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1553 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-022-669-1


  Query Match             3.5%;  Score 31.8;  DB 3;  Length 1553;
  Best Local Similarity   49.1%;  Pred. No. 2.8;
  Matches   84;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy      16 ttcaggctgtttccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaa 75
           |||||||||      |  ||| | | | ||  |||    |||||||    || || |
Db     417 TTCAGGCTGGAGGGCATTTTCCCAGCGGCGCTGCTGCCCCTGTTGCTGACCATGATTCTT 476

Qy      76 tgcctctccctgctgccgctttcctgtctgcacacgctgggaaaccggctcggacatctg 135
           | |||   ||  |||  ||    || | || |     |   |  ||| |   |     |||
Db     477 TTCCTGGGCCCACTGATGCAGCTCTCTATGGATTGCCCTTGTGACCTGGCAGATGGGCTG 536
```

```
Qy    136 gcgttttacctttaaaggaagaccgcgcgcgcatcgtcgccaatatgcgt 186
          | ||  |||             | | || |||  | || |  | ||||||
Db    537 AAGGTTGTCCTGGCCCCCCGCTCCTGGGCCCGCTGCCTCACAGACATGCGT 587


RESULT 13
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

  Query Match              3.5%;  Score 31.6;  DB 4;  Length 289;
  Best Local Similarity    8.6%;  Pred. No. 1.3;
  Matches   16;  Conservative   76;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      2 tgtttcgtttacaattcaggctgtttccccctttgcgaaccgccatgcacatcctgttga 61
          | |||  |||   |: :  |:   : : : : :|   | |:|:|     | |: :::
Db    273 TTTTTTTTTTTTTAYGYCYGYCYAYAYGYAYGYTYTYAYCYGYCYAYGYCYTYGYSYNYN 214

Qy     62 ccgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaacc 121
          ::  : ::: :   : : ::: : :::   ::: : ::: : ::: : :::       :
Db    213 YSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 154

Qy    122 ggctcggacatctggcgttttacctttaaaggaagaccgcgcgcgcatcgtcgccaata 181
          :: :   : : ::    : : : : :   :    :: : ::: : ::  : ::  :
Db    153 YSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 94

Qy    182 tgcgt 186
          ::: :
Db     93 YSYNY 89


RESULT 14
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

  Query Match              3.5%;  Score 31.6;  DB 4;  Length 289;
  Best Local Similarity    8.6%;  Pred. No. 1.3;
  Matches   16;  Conservative   76;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      2 tgtttcgtttacaattcaggctgtttccccctttgcgaaccgccatgcacatcctgttga 61
          | |||  |||   |: :  |:   : : : : :|   | |:|:|     | |: :::
Db    273 TTTTTTTTTTTTTAYGYCYGYCYAYAYGYAYGYTYTYAYCYGYCYAYGYCYTYGYSYNYN 214

Qy     62 ccgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaacc 121
          ::  : ::: :   : : ::: : :::   ::: : ::: : ::: : :::       :
Db    213 YSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 154

Qy    122 ggctcggacatctggcgttttacctttaaaggaagaccgcgcgcgcatcgtcgccaata 181
          :: :   : : ::    : : : : :   :    :: : ::: : ::  : ::  :
Db    153 YSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 94

Qy    182 tgcgt 186
          ::: :
Db     93 YSYNY 89


RESULT 15
US-08-905-223-23
; Sequence 23, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste D.
;    APPLICANT:  Duelert, Aymeric
;    APPLICANT:  Lacroix, Bruno
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,223
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
```

```
;       TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  496 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Cancerous prostate
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  149..331
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 98
;      OTHER INFORMATION:  region 1..183
;      OTHER INFORMATION:  id AA397994
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  328..485
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 96
;      OTHER INFORMATION:  region 179..336
;      OTHER INFORMATION:  id AA397994
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(182..496)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 97
;      OTHER INFORMATION:  region 14..328
;      OTHER INFORMATION:  id AA399680
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  196..240
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 5.5
;      OTHER INFORMATION:  seq ILSTVTALTFAXA/LD
US-08-905-223-23


  Query Match             3.5%;  Score 31.6;  DB 4;  Length 496;
  Best Local Similarity   50.7%;  Pred. No. 1.7;
  Matches   73;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     591 gcccgaccacgtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcgg 650
           |||||   |     ||  || ||| ||| ||| |||| |  | |    ||| |   |||
Db     120 GCCCGGAGATAGGACCAACCGTCAGGAATGCGAGGAATGTTTTTCTTCGGACTCTATCGA 179

Qy     651 caaacctgcctataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaac 710
             ||       | ||||||  | |   | | | | || ||  | | | |    |  |:| |
Db     180 GGCACACAGACAGACCATGGGGATTCTGTCTACAGTGACAGCCTTAACATTTGCCARAGC 239

Qy     711 cctgtttttctgctgcgaacgcct 734
           ||||     |||| |  |  || |
Db     240 CCTGGACGGCTGCAGAAATGGCAT 263


RESULT  16
US-09-247-155-23
; Sequence 23, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 23
;   LENGTH: 496
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 149..331
;   OTHER INFORMATION: blastn
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 328..485
;   OTHER INFORMATION: blastn
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: complement(182..496)
;   OTHER INFORMATION: blastn
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 196..240
;   OTHER INFORMATION: Von Heijne matrix
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 101
;   OTHER INFORMATION: n=a, g, c or t
US-09-247-155-23


  Query Match             3.5%;  Score 31.6;  DB 4;  Length 496;
  Best Local Similarity   50.7%;  Pred. No. 1.7;
  Matches   73;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     591 gcccgaccacgtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcgg 650
           |||||   |     ||  || ||| ||| ||| |||| |  | |    ||| |   |||
Db     120 gcccggagataggaccaaccgtcaggaatgcgaggaatgtttttcttcggactctatcga 179

Qy     651 caaacctgcctataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaac 710
             ||       | ||||||  | |   | | | | || ||  | | | |    |  |:| |
Db     180 ggcacacagacagaccatggggattctgtctacagtgacagccttaacatttgccaragc 239

Qy     711 cctgtttttctgctgcgaacgcct 734
           ||||     |||| |  |  || |
Db     240 cctggacggctgcagaaatggcat 263


RESULT  17
US-08-987-943-2
; Sequence 2, Application US/08987943
; Patent No. 6210967
; GENERAL INFORMATION:
;  APPLICANT: Bard, Jonathan A.
;  TITLE OF INVENTION: DNA Encoding A Mammalian LPA Receptor and Uses Thereof
;  FILE REFERENCE: 55183
;  CURRENT APPLICATION NUMBER: US/08/987,943
;  CURRENT FILING DATE: 1997-12-10
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 2
;   LENGTH: 1119
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-987-943-2
```

```
  Query Match             3.5%;  Score 31.4;  DB 4;  Length 1119;
  Best Local Similarity   61.7%;  Pred. No. 3.1;
  Matches   50;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      28 ccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaatgcctctccctg 87
           ||  || |||| | ||||| || | ||| || || ||||| |   | | |  | ||
Db     187 ccggctatgcgctcggccatcaacctgctgctggccaccctggccttctccgacatcatg 246

Qy      88 ctgccgctttcctgtctgcac 108
           ||| | || | |||  ||| |
Db     247 ctgtccctctgctgcatgccc 267


RESULT  18
US-08-987-943-1
; Sequence 1, Application US/08987943
; Patent No. 6210967
; GENERAL INFORMATION:
;  APPLICANT: Bard, Jonathan A.
;  TITLE OF INVENTION: DNA Encoding A Mammalian LPA Receptor and Uses Thereof
;  FILE REFERENCE: 55183
;  CURRENT APPLICATION NUMBER: US/08/987,943
;  CURRENT FILING DATE: 1997-12-10
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 1
;   LENGTH: 1755
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-987-943-1


  Query Match             3.5%;  Score 31.4;  DB 4;  Length 1755;
  Best Local Similarity   61.7%;  Pred. No. 4;
  Matches   50;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      28 ccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaatgcctctccctg 87
           ||  || |||| | ||||| || | ||| || || ||||| |   | | |  | ||
Db     536 ccggctatgcgctcggccatcaacctgctgctggccaccctggccttctccgacatcatg 595

Qy      88 ctgccgctttcctgtctgcac 108
           ||| | || | |||  ||| |
Db     596 ctgtccctctgctgcatgccc 616


RESULT  19
US-08-920-812-15/c
; Sequence 15, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5247 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-17
US-08-920-812-15


  Query Match             3.5%;  Score 31.4;  DB 1;  Length 5247;
  Best Local Similarity   64.4%;  Pred. No. 7.2;
  Matches   47;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      42 cgccatgcacatcctgttgaccgccctgctcaaatgcctctccctgctgccgctttcctg 101
           ||||||| || ||   | ||||||||||||||    | ||  |||| |  |||| | |
Db     698 CGCCATGTACTTCGGCTGGACCGCCCTGCTCATGACCATCCTCCTGGTCGCGCTGTTCCT 639

Qy     102 tctgcacacgctg 114
               |||||||
Db     638 CGGCGCCACGCTG 626


RESULT  20
US-08-920-827-15/c
; Sequence 15, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
```

```
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5247 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Pseudomonas aeruginosa
;       STRAIN:  Clinical Isolate P2-17
US-08-920-827-15


  Query Match             3.5%;  Score 31.4;  DB 1;  Length 5247;
  Best Local Similarity   64.4%;  Pred. No. 7.2;
  Matches   47;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      42 cgccatgcacatcctgttgaccgccctgctcaaatgcctctccctgctgccgctttcctg 101
           ||||||| || ||   | ||||||||||||||    | ||  |||| |  |||| | |
Db     698 CGCCATGTACTTCGGCTGGACCGCCCTGCTCATGACCATCCTCCTGGTCGCGCTGTTCCT 639

Qy     102 tctgcacacgctg 114
           |||||||
Db     638 CGGCGCCACGCTG 626


RESULT  21
US-08-921-177-15/c
; Sequence 15, Application US/08921177
; Patent No. 5798211
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,177
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5247 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-17
US-08-921-177-15


  Query Match             3.5%;  Score 31.4;  DB 1;  Length 5247;
  Best Local Similarity   64.4%;  Pred. No. 7.2;
  Matches   47;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      42 cgccatgcacatcctgttgaccgccctgctcaaatgcctctccctgctgccgctttcctg 101
           ||||||| || ||   | ||||||||||||||    | ||  |||| |  |||| | |
Db     698 CGCCATGTACTTCGGCTGGACCGCCCTGCTCATGACCATCCTCCTGGTCGCGCTGTTCCT 639

Qy     102 tctgcacacgctg 114
           |||||||
Db     638 CGGCGCCACGCTG 626


RESULT  22
US-08-362-577C-15/c
; Sequence 15, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,577C
;      FILING DATE:  27-MAR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5247 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-17
US-08-362-577C-15


  Query Match             3.5%;  Score 31.4;  DB 1;  Length 5247;
  Best Local Similarity   64.4%;  Pred. No. 7.2;
  Matches   47;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      42 cgccatgcacatcctgttgaccgccctgctcaaatgcctctccctgctgccgctttcctg 101
           ||||||| || ||   | ||||||||||||||    | ||  |||| |  |||| | |
Db     698 CGCCATGTACTTCGGCTGGACCGCCCTGCTCATGACCATCCTCCTGGTCGCGCTGTTCCT 639
```

```
Qy    102 tctgcacacgctg 114
              |||||||
Db    638 CGGCGCCACGCTG 626


RESULT 23
US-08-920-828-15/c
; Sequence 15, Application US/08920828
; Patent No. 5853998
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,828
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5247 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-17
US-08-920-828-15


  Query Match             3.5%;  Score 31.4;  DB 2;  Length 5247;
  Best Local Similarity   64.4%;  Pred. No. 7.2;
  Matches   47;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy     42 cgccatgcacatcctgttgaccgccctgctcaaatgcctctccctgctgccgctttcctg 101
          ||||||| || ||   | ||||||||||||||    | ||  |||| |  |||| | |
Db    698 CGCCATGTACTTCGGCTGGACCGCCCTGCTCATGACCATCCTCCTGGTCGCGCTGTTCCT 639

Qy    102 tctgcacacgctg 114
              |||||||
Db    638 CGGCGCCACGCTG 626


RESULT 24
US-09-134-246-8/c
; Sequence 8, Application US/09134246B
; Patent No. 6207377
; GENERAL INFORMATION:
;  APPLICANT: Wayne, Jay
;  APPLICANT: Xu, Shuang-yong
;  TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
;  TITLE OF INVENTION: Vectors And Identification of Two Thermus Plasmid
;  TITLE OF INVENTION: Replication Origins
;  FILE REFERENCE: Thermus Shuttle Vector
;  CURRENT APPLICATION NUMBER: US/09/134,246B
;  CURRENT FILING DATE: 1998-08-14
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 11958
;   TYPE: DNA
;   ORGANISM: Thermus sp.
US-09-134-246-8


  Query Match             3.5%;  Score 31.2;  DB 4;  Length 11958;
  Best Local Similarity   53.2%;  Pred. No. 13;
  Matches   66;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     43 gccatgcacatcctgttgaccgccctgctcaaatgcctctccctgctgccgctttcctgt 102
          |||| |  |||| ||  | ||  || |  | ||   |||  ||| | ||  |   | | |
Db   5731 GCCAAGGCCATCTTGGGGCCCCACCAGGGCCAAAAACTCCTCCTCCAGCTCCCGCCTTTT 5672

Qy    103 ctgcacacgctgggaaaccggctcggacatctggcgttttacctttaaaggaagaccgc 162
          |||  |  ||| ||   ||   | |  |  | | | |  ||| |||| | | |   ||   |
Db   5671 CTGGGCCAGCTTGGCCTCCCCGTAGCGCTCCAGCCACTTTTCCTTGTCAGGCCGGAGGTC 5612

Qy    163 gcgc 166
          || |
Db   5611 GCAC 5608


RESULT 25
US-08-814-052-15
; Sequence 15, Application US/08814052
; Patent No. 6015783
;  GENERAL INFORMATION:
;    APPLICANT:  von der Osten, Claus
;    APPLICANT:  Cherry, Joel R.
;    APPLICANT:  Bjornvad, Mads E.
;    APPLICANT:  Vind, Jesper
;    APPLICANT:  Rasmussen, Michael Dolberg
;    TITLE OF INVENTION:  PROCESS FOR REMOVAL OR BLEACHING OF SOILING
;    TITLE OF INVENTION:  OR STAINS FROM CELLULOSIC FABRIC
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6015783o No. 6015783disk of No. 6015783th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/814,052
;      FILING DATE:  06-MAR-1997
;      CLASSIFICATION:  510
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4684.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;      TELEX:
```

```
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  480 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-814-052-15


  Query Match             3.5%;  Score 31;  DB 3;  Length 480;
  Best Local Similarity   56.3%;  Pred. No. 2.6;
  Matches   58;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    565 cgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggcggg 624
          | |  | ||   ||  || ||||   |||| || || |||| |||| ||||     ||
Db    219 CATCAGTGCCTCGCCCCCGTCGTCGCCCCCGCCCCCGCCCCCGCCCCCCAAGGTCCTGGA 278

Qy    625 gaaggcgtatgggtggatttcttcggcaaacctgcctatacca 667
          | ||||| |  ||  || |  |||   ||   ||  | | |
Db    279 GGAGGCGGGTCAGTCACTTGCCCCGGTGGACAGTCCACTTCGA 321


RESULT  26
US-08-715-252-1/c
; Sequence 1, Application US/08715252
; Patent No. 5726043
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Process for the production of protease
;    TITLE OF INVENTION:  inhibitors
;    NUMBER OF SEQUENCES:  7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/715,252
;      FILING DATE:  16-SEP-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/116,988
;      FILING DATE:  30-AUG-1993
;      APPLICATION NUMBER:  EP 92810681.4
;      FILING DATE:  04-SEP-1992
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1082 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..6
;      OTHER INFORMATION:  /function= "BamHI linker"
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  7..432
;      OTHER INFORMATION:  /standard_name= "CUP1 promoter"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  433..441
;      OTHER INFORMATION:  /function= "EcoRI linker"
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  442..492
;      OTHER INFORMATION:  /standard_name= "PHO5 signal
;      OTHER INFORMATION:  sequence"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  493..690
;      OTHER INFORMATION:  /product= "Desulphatohirudin HV1"
;      OTHER INFORMATION:  /standard_name= "HV1"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  691..1068
;      OTHER INFORMATION:  /standard_name= "PHO5 transcription
;      OTHER INFORMATION:  terminator"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1069..1082
;      OTHER INFORMATION:  /function= "SalI linker"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  442..690
;      OTHER INFORMATION:  /product= "primary transcript"
US-08-715-252-1


  Query Match             3.5%;  Score 31;  DB 1;  Length 1082;
  Best Local Similarity   57.9%;  Pred. No. 4.1;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy    696 caaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcctggcggacaaggtttcga 755
          || || |||  ||  | |   ||||    | |||| |  | |  | |||  ||||||| |
Db    852 CATAGTCGTAGAAGTCATTGATTTCACAAGAGAACCCTGGACCAGTGGAACAGGTTTCAA 793

Qy    756 tttgcacatccgccccgtccaaggggaattgaacg 790
          || | ||| | ||  |||  | |  | ||   |||
Db    792 TTGGAACAACAGCATCGTTAATGACGTATCTGACG 758


RESULT  27
US-08-453-051-3/c
; Sequence 3, Application US/08453051
; Patent No. 5922569
;  GENERAL INFORMATION:
;    APPLICANT:  Heim, Jutta
;    APPLICANT:  Hottiger, Thomas
;    APPLICANT:  Pohlig, Gabriele
;    APPLICANT:  Furst, Peter
;    TITLE OF INVENTION:  Process for the production of peptides
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  PO Box 12257
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709-2257
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,051
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/165,720
;      FILING DATE:
;      APPLICATION NUMBER:  GB 92811005.5
;      FILING DATE:  12-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  4-19393/A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1082 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..6
;       OTHER INFORMATION:  /function= "BamHI linker"
;     FEATURE:
;       NAME/KEY:  promoter
;       LOCATION:  7..432
;       OTHER INFORMATION:  /standard_name= "CUP1 promoter"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  433..441
;       OTHER INFORMATION:  /function= "EcoRI linker"
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  442..492
;       OTHER INFORMATION:  /standard_name= "PHO5 signal
;       OTHER INFORMATION:  sequence"
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  493..690
;       OTHER INFORMATION:  /product= "desulphatohirudin"
;       OTHER INFORMATION:  /standard_name= "HV1"
;     FEATURE:
;       NAME/KEY:  terminator
;       LOCATION:  691..1068
;       OTHER INFORMATION:  /standard_name= "PHO5 terminator"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1069..1082
;       OTHER INFORMATION:  /function= "SalI linker"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  442..690
;       OTHER INFORMATION:  /product= "primary transcript"
US-08-453-051-3


  Query Match             3.5%;  Score 31;  DB 2;  Length 1082;
  Best Local Similarity   57.9%;  Pred. No. 4.1;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     696 caaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcctggcggacaaggtttcga 755
           || || |||  ||  | |   ||||    | |||| |  | |  | |||  ||||||| |
Db     852 CATAGTCGTAGAAGTCATTGATTTCACAAGAGAACCCTGGACCAGTGGAACAGGTTTCAA 793

Qy     756 tttgcacatccgccccgtccaaggggaattgaacg 790
           || | ||| | ||  |||  | |  | ||   |||
Db     792 TTGGAACAACAGCATCGTTAATGACGTATCTGACG 758


RESULT  28
US-08-406-948A-5/c
; Sequence 5, Application US/08406948A
; Patent No. 5661001
;  GENERAL INFORMATION:
;    APPLICANT:  Grossenbacher, Hugh
;    APPLICANT:  Jui Yoa, Chang
;    APPLICANT:  Walter, M rki
;    TITLE OF INVENTION:  High Molecular Weight Hirudin
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      ZIP:  10591-9005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/406,948A
;      FILING DATE:  March 29, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/02438
;      FILING DATE:  23 JUL 94
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferraro, Gregory D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  4-19636/A/PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 277-3318
;      TELEFAX:  (908) 277-4306
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1130 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  1..1130
;      OTHER INFORMATION:  /function= "first 1130 nucleotides
;      OTHER INFORMATION:  of pJDB207/GAPFL-YHIR (EP-A-340 170)"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..276
;      OTHER INFORMATION:  /product= "SalI - BamHI fragment of
;      OTHER INFORMATION:  pBR322"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  277..284
;      OTHER INFORMATION:  /function= "BamHI/BglII linker"
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  285..478
;      OTHER INFORMATION:  /function= "GAPFL promoter"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  479..484
;      OTHER INFORMATION:  /product= "EcoRI linker"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  488..538
;      OTHER INFORMATION:  /function= "PHO5 signal sequence"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  539..733
;      OTHER INFORMATION:  /product= "Hirudin HV1"
;      OTHER INFORMATION:  /standard_name= "HV1"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  737..1113
;      OTHER INFORMATION:  /standard_name= "PHO5 terminator"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1117..1122
;      OTHER INFORMATION:  /function= "M13 cloning site"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  734..736
;      OTHER INFORMATION:  /function= "Translation terminator"
;      OTHER INFORMATION:  /standard_name= "TAG"
US-08-406-948A-5


  Query Match             3.5%;  Score 31;  DB 1;  Length 1130;
  Best Local Similarity   57.9%;  Pred. No. 4.2;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;
```

```
Qy    696 caaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcctggcggacaaggtttcga 755
          || || ||| || | |  |||| | |||| | | | ||| |||||||
Db    898 CATAGTCGTAGAAGTCATTGATTTCACAAGAGAACCCTGGACCAGTGGAACAGGTTTCAA 839

Qy    756 tttgcacatccgccccgtccaaggggaattgaacg 790
          || | ||| | ||  ||| | | | ||  |||
Db    838 TTGGAACAACAGCATCGTTAATGACGTATCTGACG 804


RESULT 29
US-09-248-571-2
; Sequence 2, Application US/09248571
; Patent No. 6136539
; GENERAL INFORMATION:
;  APPLICANT: BASBAUM, CAROL
;  APPLICANT: GALLUP, MARIANNE
;  APPLICANT: DAIZONG, LI
;  APPLICANT: GEBREMICHAEL, ASSEFA
;  APPLICANT: GENSCH, ERIN
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITION OF MUC-5 MUCIN
;  TITLE OF INVENTION: GENE EXPRESSION
;  FILE REFERENCE: UCSF12/02
;  CURRENT APPLICATION NUMBER: US/09/248,571
;  CURRENT FILING DATE: 1999-02-11
;  EARLIER APPLICATION NUMBER: 60/074,398
;  EARLIER FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3358
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-248-571-2


  Query Match             3.3%;  Score 30;  DB 3;  Length 3358;
  Best Local Similarity   59.3%;  Pred. No. 16;
  Matches   51;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    345 ggacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgg 404
          |||| |   | | | | ||| || || ||  | ||| || |||| ||||    |||  |
Db   3002 ggacgagagccaggaggtgccatacaccatccggcagatgggcatctacctggtggtgga 3061

Qy    405 acgctacatcagccagcagcttccgt 430
            | ||||  ||| |  ||| | ||
Db   3062 caccgacattggcctggtgctgctgt 3087


RESULT 30
US-09-404-650-1
; Sequence 1, Application US/09404650
; Patent No. 6309858
; GENERAL INFORMATION:
;  APPLICANT: Dietrich, Paul S.
;  APPLICANT: McGivern, Joseph G.
;  TITLE OF INVENTION: T-TYPE CALCIUM CHANNEL VARIANTS; COMPOSITIONS THEREOF;
;  TITLE OF INVENTION: AND USES
;  FILE REFERENCE: R0043B-REG sequence listing
;  CURRENT APPLICATION NUMBER: US/09/404,650
;  CURRENT FILING DATE: 1999-09-23
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 6816
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (192)..(6716)
US-09-404-650-1


  Query Match             3.3%;  Score 30;  DB 4;  Length 6816;
  Best Local Similarity   55.9%;  Pred. No. 23;
  Matches   57;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    346 gacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgga 405
          |||||     | ||| |  |      ||||| ||    | ||| | ||||   | || |
Db   5118 gacaactggaacgggatcatgaaggacacgctgcgggactgcacccacgacgagcgcagc 5177

Qy    406 cgctacatcagccagcagcttccgttcccgctgaccgccatg 447
            ||   | ||||| |||| ||  ||  ||||||  |  | ||
Db   5178 tgcctgagcagcctgcagtttgtgtcgccgctgtacttcgtg 5219


RESULT 31
US-09-404-650-3
; Sequence 3, Application US/09404650
; Patent No. 6309858
; GENERAL INFORMATION:
;  APPLICANT: Dietrich, Paul S.
;  APPLICANT: McGivern, Joseph G.
;  TITLE OF INVENTION: T-TYPE CALCIUM CHANNEL VARIANTS; COMPOSITIONS THEREOF;
;  TITLE OF INVENTION: AND USES
;  FILE REFERENCE: R0043B-REG sequence listing
;  CURRENT APPLICATION NUMBER: US/09/404,650
;  CURRENT FILING DATE: 1999-09-23
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 6855
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (192)..(6755)
US-09-404-650-3


  Query Match             3.3%;  Score 30;  DB 4;  Length 6855;
  Best Local Similarity   55.9%;  Pred. No. 23;
  Matches   57;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    346 gacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgga 405
          |||||     | ||| |  |      ||||| ||    | ||| | ||||   | || |
Db   5118 gacaactggaacgggatcatgaaggacacgctgcgggactgcacccacgacgagcgcagc 5177

Qy    406 cgctacatcagccagcagcttccgttcccgctgaccgccatg 447
            ||   | ||||| |||| ||  ||  ||||||  |  | ||
Db   5178 tgcctgagcagcctgcagtttgtgtcgccgctgtacttcgtg 5219


RESULT 32
US-08-481-626-1/c
; Sequence 1, Application US/08481626
; Patent No. 5801040
;  GENERAL INFORMATION:
;    APPLICANT:  Soubrier, Florent
;    APPLICANT:  Alhenc-Gelas, Francois
;    APPLICANT:  Hubert, Christine
;    APPLICANT:  Corvol, Pierre
;    TITLE OF INVENTION:  Nucleic Acid Coding for the Human
;    TITLE OF INVENTION:  Testicular Angiotensin Converting Enzyme (ACE) and its
;    TITLE OF INVENTION:  Uses, Especially for the In Vitro Screening for this
;    TITLE OF INVENTION:  Enzyme in the Organism
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/481,626
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/656,183
;       FILING DATE:  04-MAR-1991
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 89-09062
;       FILING DATE:  05-JUL-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meyers, Kenneth J.
;       REGISTRATION NUMBER:  25,146
;       REFERENCE/DOCKET NUMBER:  04958-0006-02000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-408-4000
;       TELEFAX:  202-408-4400
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2478 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-481-626-1


  Query Match             3.3%;  Score 29.8;  DB 1;  Length 2478;
  Best Local Similarity   47.6%;  Pred. No. 16;
  Matches   88;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      88 ctgccgctttcctgtctgcacacgctgggaaaccggctcggacatctggcgttttacctt 147
           | ||  | |  | |||  ||  |||||     ||  || ||||   | |   |    |||
Db    1825 CAGCTTCATGGCGGTCGCCAGGCGCTGCCCGGCCTCCTTGGACTGGTAGATGTCACACTT 1766

Qy     148 ttaaaggaagaccgcgcgcgcatcgtcgccaatatgcgtcaggcaggcatgaatcccgac 207
             |  |||  | ||| | |  || |  |         | | |  |  |   || | |  ||
Db    1765 GTGCAGGGGGCCCGTGTGGCCAGCTGCCTGGCACAGTGCCTCGTGGAACTGGAACTGGAT 1706

Qy     208 cccaaaacggtcaaagccgtttttgcggaaacggcaaaaggcggtttggaacttgccccc 267
              || | | |||||    |  ||        || | | |    | |||||||| ||||
Db    1705 GATGAAGCTGACAAAGTACCTGATGTAAGGCACGCTAGAAGGAATGTGGAACTTGGCCCC 1646

Qy     268 gcgtt 272
             | |
Db    1645 TGGGT 1641


RESULT  33
US-09-050-159-130/c
; Sequence 130, Application US/09050159A
; Patent No. 6197505
; GENERAL INFORMATION:
;  APPLICANT: No. 6197505berg, Leif T
;  APPLICANT: Andersson, Maria K
;  APPLICANT: Linstrom, Per H
;  TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND
;  TITLE OF INVENTION: COMPOSITIONS FOR USE THEREOF
;  FILE REFERENCE: 1248/1D042
;  CURRENT APPLICATION NUMBER: US/09/050,159A
;  CURRENT FILING DATE: 1998-03-27
;  EARLIER APPLICATION NUMBER: 60/042,930
;  EARLIER FILING DATE: 1987-04-03
;  NUMBER OF SEQ ID NOS: 133
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 130
;   LENGTH: 4020
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Angiotensin I converting enzyme mRNA
US-09-050-159-130


  Query Match             3.3%;  Score 29.8;  DB 4;  Length 4020;
  Best Local Similarity   47.6%;  Pred. No. 20;
  Matches   88;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      88 ctgccgctttcctgtctgcacacgctgggaaaccggctcggacatctggcgttttacctt 147
           | ||  | |  | |||  ||  |||||     ||  || ||||   | |   |    |||
Db    3541 CAGCTTCATGGCGGTCGCCAGGCGCTGCCCGGCCTCCTTGGACTGGTAGATGTCACACTT 3482

Qy     148 ttaaaggaagaccgcgcgcgcatcgtcgccaatatgcgtcaggcaggcatgaatcccgac 207
             |  |||  | ||| | |  || |  |         | | |  |  |   || | |  ||
Db    3481 GTGCAGGGGGCCCGTGTGGCCAGCTGCCTGGCACAGTGCCTCGTGGAACTGGAACTGGAT 3422

Qy     208 cccaaaacggtcaaagccgtttttgcggaaacggcaaaaggcggtttggaacttgccccc 267
              || | | |||||    |  ||        || | | |    | |||||||| ||||
Db    3421 GATGAAGCTGACAAAGTACCTGATGTAAGGCACGCTAGAAGGAATGTGGAACTTGGCCCC 3362

Qy     268 gcgtt 272
             | |
Db    3361 TGGGT 3357


RESULT  34
US-09-162-484-18/c
; Sequence 18, Application US/09162484
; Patent No. 6248724
; GENERAL INFORMATION:
;  APPLICANT: Phillips, M. Ian
;  APPLICANT: Mohuczy, Dagmara
;  TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE COMPOSITIONS TARGETED TO
;  TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME MRNA AND METHODS OF USE
;  FILE REFERENCE: UFLA:087/UFLA087P
;  CURRENT APPLICATION NUMBER: US/09/162,484
;  CURRENT FILING DATE: 1998-09-25
;  EARLIER APPLICATION NUMBER: 60/059,661
;  EARLIER FILING DATE: 1997-09-25
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 4024
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-162-484-18


  Query Match             3.3%;  Score 29.8;  DB 4;  Length 4024;
  Best Local Similarity   47.6%;  Pred. No. 20;
  Matches   88;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      88 ctgccgctttcctgtctgcacacgctgggaaaccggctcggacatctggcgttttacctt 147
           | ||  | |  | |||  ||  |||||     ||  || ||||   | |   |    |||
Db    3541 CAGCTTCATGGCGGTCGCCAGGCGCTGCCCGGCCTCCTTGGACTGGTAGATGTCACACTT 3482

Qy     148 ttaaaggaagaccgcgcgcgcatcgtcgccaatatgcgtcaggcaggcatgaatcccgac 207
             |  |||  | ||| | |  || |  |         | | |  |  |   || | |  ||
Db    3481 GTGCAGGGGGCCCGTGTGGCCAGCTGCCTGGCACAGTGCCTCGTGGAACTGGAACTGGAT 3422

Qy     208 cccaaaacggtcaaagccgtttttgcggaaacggcaaaaggcggtttggaacttgccccc 267
              || | | |||||    |  ||        || | | |    | |||||||| ||||
Db    3421 GATGAAGCTGACAAAGTACCTGATGTAAGGCACGCTAGAAGGAATGTGGAACTTGGCCCC 3362

Qy     268 gcgtt 272
             | |
Db    3361 TGGGT 3357


RESULT  35
US-07-794-393-1/c
; Sequence 1, Application US/07794393
; Patent No. 5236844
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, PIERRE
;    APPLICANT:  BASSET, PAUL
;    APPLICANT:  BELLOCQ, JEAN-PIERRE
;    TITLE OF INVENTION:  ANALYTICAL MARKERS FOR MALIGNANT BREAST
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1225 Connecticut Ave. NW Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/794,393
;      FILING DATE:  19911121
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9025326.1
;      FILING DATE:  21-NOV-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GOLDSTEIN, JORGE A
;      REGISTRATION NUMBER:  29,021
;      REFERENCE/DOCKET NUMBER:  1383.0040000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 466-0800
;      TELEFAX:  (202) 833-8716
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2256 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..1473
US-07-794-393-1


  Query Match             3.3%;  Score 29.6;  DB 1;  Length 2256;
  Best Local Similarity   44.8%;  Pred. No. 17;
  Matches  113;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy    346 gacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgga 405
          ||  ||    |  ||   |||   ||    |     |  || | | | ||||   ||| | |
Db   1473 GAGGAAAGTGTTGGCAGGCTCGGCACAGCCAAAGAAGTCAGGACCCACGAGACGGGGGAA 1414

Qy    406 cgctacatcagccagcagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatc 465
            || |   ||||  || ||||                |  |  |     |  || |
Db   1413 GCCTTCCAGAGCCTTCACCTTCACAGGGTCAAACTTCCAGTAGAGGCGGCCGCGCAGGAA 1354

Qy    466 aaagcgatagacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcct 525
            ||  |||| ||  | ||| | ||  ||||| ||      || |||  |  || | ||
Db   1353 GTAGGCATAGCCATCAGCATCCTGGAAGGCAGCGTCGATCTCAGAGGGCACCCCTCTCCA 1294

Qy    526 accagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatc 585
            |||    |  | ||   |    |   || |  | | | || | || |  | |||
Db   1293 GTCAGTGGCCCTGCGGGGCACGGGACTGTCTACACGCCGGGTGCTGGGGTGGAAACGCCA 1234

Qy    586 gtcctgcccgac 597
            ||  | || | |
Db   1233 GTAGTCCCTGCC 1222


RESULT  36
US-08-001-711-1/c
```

```
; Sequence 1, Application US/08001711
; Patent No. 5484726
;  GENERAL INFORMATION:
;    APPLICANT:  BASSET, PAUL
;    APPLICANT:  BELLOCQ, JEAN-PIERRE
;    APPLICANT:  CHAMBON, PIERRE
;    TITLE OF INVENTION:  ANALYTICAL MARKERS FOR MALIGNANT BREAST
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1225 Connecticut Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/001,711
;      FILING DATE:  19930107
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/794,393
;      FILING DATE:  11-NOV-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9025626.1
;      FILING DATE:  21-NOV-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MILLMAN, ROBERT A
;      REGISTRATION NUMBER:  36,217
;      REFERENCE/DOCKET NUMBER:  1383.0040001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)466-0800
;      TELEFAX:  (202)833-8716
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2256 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..1473
US-08-001-711-1


  Query Match             3.3%;  Score 29.6;  DB 1;  Length 2256;
  Best Local Similarity   44.8%;  Pred. No. 17;
  Matches  113;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy    346 gacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgga 405
          ||  ||    |  ||   |||   ||    |     |  || | | | ||||   ||| | |
Db   1473 GAGGAAAGTGTTGGCAGGCTCGGCACAGCCAAAGAAGTCAGGACCCACGAGACGGGGGAA 1414

Qy    406 cgctacatcagccagcagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatc 465
            || |   ||||  || ||||                |  |  |     |  || |
Db   1413 GCCTTCCAGAGCCTTCACCTTCACAGGGTCAAACTTCCAGTAGAGGCGGCCGCGCAGGAA 1354

Qy    466 aaagcgatagacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcct 525
            ||  |||| ||  | ||| | ||  ||||| ||      || |||  |  || | ||
Db   1353 GTAGGCATAGCCATCAGCATCCTGGAAGGCAGCGTCGATCTCAGAGGGCACCCCTCTCCA 1294

Qy    526 accagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatc 585
            |||    |  | ||   |    |   || |  | | | || | || |  | |||
Db   1293 GTCAGTGGCCCTGCGGGGCACGGGACTGTCTACACGCCGGGTGCTGGGGTGGAAACGCCA 1234

Qy    586 gtcctgcccgac 597
            ||  | || | |
Db   1233 GTAGTCCCTGCC 1222
```

```
RESULT 37
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-07-642-734C-1


  Query Match             3.3%;  Score 29.4;  DB 1;  Length 11219;
  Best Local Similarity   51.1%;  Pred. No. 48;
  Matches   69;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     570 gggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggcggggaagg 629
           | |||| |  | ||||| ||||| | || ||  |||  | || |  |    || |||||
Db    2129 GCGCGAGGTCAACATCGCCCTGCTCTACGACCACCCGACTCCGCGTGCGCTCGCGGAAGC 2188

Qy     630 cgtatgggtggatttcttcggcaaacctgcctataccatgacgctggcggcaaaattggc 689
             |   || ||    |   | |  ||      | |||    ||| | |     ||  |||
Db    2189 ACTCGCGGCGGGAACCGAGGTCGCACAACGGGAAACCCGCGCGCGGACCAACGAAGCGGC 2248

Qy     690 acacgtcaaaggcgt 704
           | || | ||   ||
Db    2249 GCCCGGCGAACCGGT 2263


RESULT 38
US-08-439-009A-1
; Sequence 1, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
```

```
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-08-439-009A-1


  Query Match             3.3%;  Score 29.4;  DB 3;  Length 11219;
  Best Local Similarity   51.1%;  Pred. No. 48;
  Matches   69;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy    570 gggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggcggggaagg 629
          | |||| |  | ||||| ||||| | || ||  |||  | || |  |    || |||||
Db   2129 GCGCGAGGTCAACATCGCCCTGCTCTACGACCACCCGACTCCGCGTGCGCTCGCGGAAGC 2188

Qy    630 cgtatgggtggatttcttcggcaaacctgcctataccatgacgctggcggcaaaattggc 689
            |   || ||    |   | |  ||      | |||    ||| | |     ||  |||
Db   2189 ACTCGCGGCGGGAACCGAGGTCGCACAACGGGAAACCCGCGCGCGGACCAACGAAGCGGC 2248

Qy    690 acacgtcaaaggcgt 704
            | || | ||   ||
Db   2249 GCCCGGCGAACCGGT 2263


RESULT  39
US-08-905-223-196
; Sequence 196, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste D.
;    APPLICANT:  Duelert, Aymeric
;    APPLICANT:  Lacroix, Bruno
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/905,223
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  196:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  342 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Brain
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  33..269
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 99
;      OTHER INFORMATION:  region 10..246
;      OTHER INFORMATION:  id AA058587
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  272..307
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 100
;      OTHER INFORMATION:  region 248..283
;      OTHER INFORMATION:  id AA058587
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  133..259
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 100
;      OTHER INFORMATION:  region 87..213
;      OTHER INFORMATION:  id R12128
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  47..134
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 98
;      OTHER INFORMATION:  region 2..89
;      OTHER INFORMATION:  id R12128
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  303..337
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 97
;      OTHER INFORMATION:  region 257..291
;      OTHER INFORMATION:  id R12128
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  49..259
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 98
;      OTHER INFORMATION:  region 2..212
;      OTHER INFORMATION:  id H19999
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  272..304
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 96
;      OTHER INFORMATION:  region 225..257
;      OTHER INFORMATION:  id H19999
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  303..337
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 91
;      OTHER INFORMATION:  region 257..291
;      OTHER INFORMATION:  id H19999
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  42..252
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 99
;      OTHER INFORMATION:  region 1..211
;      OTHER INFORMATION:  id R20025
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  87..259
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 99
;      OTHER INFORMATION:  region 1..173
;      OTHER INFORMATION:  id H83838
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  272..337
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 98
;      OTHER INFORMATION:  region 186..251
;      OTHER INFORMATION:  id H83838
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  85..198
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 9.3
;      OTHER INFORMATION:  seq QLLYLSLLSGLHG/QE
US-08-905-223-196


  Query Match             3.3%;  Score 29.2;  DB 4;  Length 342;
  Best Local Similarity   57.8%;  Pred. No. 8.2;
  Matches   52;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      53 tcctgttgaccgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgc 112
           || || ||   ||  ||||  |    |||| ||||  || ||| ||| | ||||||  ||
Db     140 TCATGCTGGTGGCGATGCTGCAGCTGCTCTACCTGTCGCTGCTGTCCGGACTGCACGGGC 199

Qy     113 tgggaaaccggctcggacatctggcgtttt 142
             ||   | |   | | || | | ||| |
Db     200 AGGAGGAGCAAGACCAATATTTTGAGTTCT 229


RESULT  40
US-07-914-282D-3
; Sequence 3, Application US/07914282D
; Patent No. 5364787
;  GENERAL INFORMATION:
;    APPLICANT:  Orser, Cindy S. and Xun, Luying
;    TITLE OF INVENTION:  GENES AND ENZYMES INVOLVED IN
;    TITLE OF INVENTION:  THE MICROBIAL DEGRADATION OF
;    TITLE OF INVENTION:  PENTACHLOROPHENOL
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ramon A. Klitzke II
;      STREET:  One World Trade Center
;      STREET:  121 S.W. Salmon Street
;      STREET:  Suite 1600
;      CITY:  Portland
```

```
;       STATE:  Oregon
;       COUNTRY:  United States of America
;       ZIP:  97204
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3-1/2 inch
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  MS DOS
;       SOFTWARE:  WordPerfect 5.1/PC Gene
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/914,282D
;       FILING DATE:  July 13, 1992
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/856,015
;       FILING DATE:  March 23, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Ramon A. Klitzke II
;       REGISTRATION NUMBER:  30,188
;       REFERENCE/DOCKET NUMBER:  2815-36746
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (503) 226-7391
;       TELEFAX:  (503) 228-9446
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1432 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Double-stranded
;       TOPOLOGY:  Linear
;     MOLECULE TYPE:  Genomic DNA
;       DESCRIPTION:
;     HYPOTHETICAL:  No
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Flavobacterium sp. Strain ATCC 39723
US-07-914-282D-3


  Query Match             3.3%;  Score 29.2;  DB 1;  Length 1432;
  Best Local Similarity   49.4%;  Pred. No. 18;
  Matches   76;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy     306 caaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaagggctgct 365
           ||  | ||  || |     |   | |||| ||    | ||| |||  || | ||  ||
Db     879 CACGGTGGCCCAGGCGGAAGTCGACCTGCAAAAGCGCCTCGACGAACTGGACGTGCATCT 938

Qy     366 attcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagct 425
               || ||||  |||||   ||  | | |   ||     | ||||||    |    |
Db     939 GGCGGACAAGCCGTTCATCGCGGGCAGCAACTACAGCATCGCCGACATCATGTGGACCGT 998

Qy     426 tccgttcccgctgaccgccatgtacaaaccgccg 459
           | | |  ||| || ||  |||  |||   | ||
Db     999 CCTGCTGGCGCGGATCGAGATGCTCAACATGACG 1032


RESULT  41
US-08-276-887A-3
; Sequence 3, Application US/08276887A
; Patent No. 5512478
;  GENERAL INFORMATION:
;    APPLICANT:  Orser, Cindy S. and Xun, Luying
;    TITLE OF INVENTION:  GENES AND ENZYMES INVOLVED
;    TITLE OF INVENTION:  IN THE MICROBIAL
;    TITLE OF INVENTION:  DEGRADATION OF
;    TITLE OF INVENTION:  PENTACHLOROPHENOL
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ramon A. Klitzke II
;      STREET:  One World Trade Center
;      STREET:  121 S.W. Salmon Street
;      STREET:  Suite 1600
;      CITY:  Portland
;      STATE:  Oregon
;      COUNTRY:  United States of America
;      ZIP:  97204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3-1/2 inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1/PC Gene
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,887A
;      FILING DATE:  July 18, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/914,282
;      FILING DATE:  July 13, 1992
;      APPLICATION NUMBER:  07/856,015
;      FILING DATE:  March 23, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ramon A. Klitzke II
;      REGISTRATION NUMBER:  30,188
;      REFERENCE/DOCKET NUMBER:  2815-36746
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (503) 226-7391
;      TELEFAX:  (503) 228-9446
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1432 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Double-stranded
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  Genomic DNA
;      DESCRIPTION:
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Flavobacterium sp. Strain
;      ORGANISM:  ATCC 39723
US-08-276-887A-3


  Query Match             3.3%;  Score 29.2;  DB 1;  Length 1432;
  Best Local Similarity   49.4%;  Pred. No. 18;
  Matches   76;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy     306 caaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaagggctgct 365
           ||  | ||  || |     |   | |||| ||    | ||| |||  || | ||  ||
Db     879 CACGGTGGCCCAGGCGGAAGTCGACCTGCAAAAGCGCCTCGACGAACTGGACGTGCATCT 938

Qy     366 attcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagct 425
               || ||||  |||||   ||  | | |   ||     | ||||||    |    |
Db     939 GGCGGACAAGCCGTTCATCGCGGGCAGCAACTACAGCATCGCCGACATCATGTGGACCGT 998

Qy     426 tccgttcccgctgaccgccatgtacaaaccgccg 459
           | | |  ||| || ||  |||  |||   | ||
Db     999 CCTGCTGGCGCGGATCGAGATGCTCAACATGACG 1032


RESULT  42
PCT-US93-02460-3
; Sequence 3, Application PC/TUS9302460
;  GENERAL INFORMATION:
;    APPLICANT:  Orser, Cindy S. and Xun, Luying
;    TITLE OF INVENTION:  GENES AND ENZYMES INVOLVED IN THE
;    TITLE OF INVENTION:  MICROBIAL DEGRADATION OF PENTACHLOROPHENOL
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  William D. Noonan, M.D.
;      STREET:  One World Trade Center
;      STREET:  121 S.W. Salmon Street
;      STREET:  Suite 1600
;      CITY:  Portland
;      STATE:  Oregon
;      COUNTRY:  United States of America
```

```
;       ZIP:  97204
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5-inch
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  MS DOS
;       SOFTWARE:  WordPerfect 5.1/PC Gene
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/02460
;       FILING DATE:  19930319
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/856,015
;       FILING DATE:  March 23, 1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/914,282
;       FILING DATE:  July 13, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William D. Noonan, M.D.
;       REGISTRATION NUMBER:  30878
;       REFERENCE/DOCKET NUMBER:  2815-36746-WDN
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (503) 226-7391
;       TELEFAX:  (503) 228-9446
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1432 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  Double-stranded
;       TOPOLOGY:  Linear
;     MOLECULE TYPE:  Genomic DNA
;       DESCRIPTION:
;     HYPOTHETICAL:  No
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Flavobacterium sp. Strain ATCC 39723
PCT-US93-02460-3


  Query Match              3.3%;  Score 29.2;  DB 5;  Length 1432;
  Best Local Similarity   49.4%;  Pred. No. 18;
  Matches   76;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy     306 caaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaagggctgct 365
           ||  | ||  || |     |   |  ||||  ||    | ||| |||  || | ||  ||
Db     879 CACGGTGGCCCAGGCGGAAGTCGACCTGCAAAAGCGCCTCGACGAACTGGACGTGCATCT 938

Qy     366 attcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagct 425
               || ||||  |||||   ||  | | |   ||     | ||||||    |    |
Db     939 GGCGGACAAGCCGTTCATCGCGGGCAGCAACTACAGCATCGCCGACATCATGTGGACCGT 998

Qy     426 tccgttcccgctgaccgccatgtacaaaccgccg 459
           | | |  ||| || ||  |||  |||   | ||
Db     999 CCTGCTGGCGCGGATCGAGATGCTCAACATGACG 1032


RESULT  43
US-07-945-283-3
; Sequence 3, Application US/07945283
; Patent No. 5352596
;  GENERAL INFORMATION:
;    APPLICANT:  Cheung, Andrew K.
;    APPLICANT:  Wesley, Ronald D.
;    TITLE OF INVENTION:  Pseudorabies Virus Deletion Mutants
;    TITLE OF INVENTION:  Involving The EP0 and LLT Genes
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis P. Ribando
;      STREET:  1815 No. 5352596th University Street
;      CITY:  Peoria
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  61604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/945,283
;      FILING DATE:  19920911
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ribando, Curtis P
;      REGISTRATION NUMBER:  27976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  309-685-4011 ext.513
;      TELEFAX:  309-685-4128
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1683 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudorabies virus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..1440
;      OTHER INFORMATION:  /product= "early protein 0"
US-07-945-283-3


  Query Match              3.3%;  Score 29.2;  DB 1;  Length 1683;
  Best Local Similarity   54.7%;  Pred. No. 20;
  Matches   58;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      21 gctgtttccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaatgcct 80
           |||  | |||||  | ||   | |  | |   |||||  |  | ||||  |||   | |||
Db    1500 GCTCCTGCCCCCACTTCGGCTCCCGCTCCTGCTCCTGCTCCCGGCCCCACTCCTGTTCCT 1559

Qy      81 ctccctgctgccgctttcctgtctgcacacgctgggaaaccggctc 126
               | |||| | | | |    |  |||  | || |      |||||||
Db    1560 GTTCCTGTTCCTGTTCCTGTTCCTGTTCCCGGTCCTGCTCCGGCTC 1605


RESULT  44
US-09-321-981-4/c
; Sequence 4, Application US/09321981
; Patent No. 6287839
; GENERAL INFORMATION:
;  APPLICANT:  Genencor International, Inc.
;  TITLE OF INVENTION:  No. 6287839el Cellulase Producing Actinomycetes,
;  TITLE OF INVENTION:  Cellulase Produced Therefrom and Method of Producing Same
;  FILE REFERENCE:  GC540-2
;  CURRENT APPLICATION NUMBER: US/09/321,981
;  CURRENT FILING DATE:  1999-05-28
;  PRIOR APPLICATION NUMBER:  PCT/US99/11971
;  PRIOR FILING DATE:  1999-05-28
;  PRIOR APPLICATION NUMBER:  09/104,308
;  PRIOR FILING DATE:  1998-06-24
;  PRIOR APPLICATION NUMBER:  08/974,042
;  PRIOR FILING DATE:  1997-11-19
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 1716
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Nearest "neighbor" = Streptomycetes
;   OTHER INFORMATION: thermovebaceous
US-09-321-981-4
```

```
  Query Match             3.3%;  Score 29.2;  DB 4;  Length 1716;
  Best Local Similarity   52.5%;  Pred. No. 20;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     260 ttgcccccgcgtttttcagaaaaccggaagacatagaaacaatgttcaaagcggtacacg 319
           || | |  | | |   | | |||||| | |||   |   |   |||    ||| | |
Db    1509 TTCCGCTGGAGGTGCCCTGGAAACCGAAGGACTGGGTCCCGCCGTTGGCCGCGATCCGGC 1450

Qy     320 gctgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatcacgccgc 379
             | |||||  |||| ||  ||  ||  |   | ||| | |  |||  ||| | ||   |
Db    1449 CGTTGGAACCGGTGCTGCGTGCCGTGACCGCGCCCGAGGCGGGGCTGATCAGCGCGTTCC 1390

Qy     380 ac 381
           ||
Db    1389 AC 1388


RESULT  45
US-08-714-168-3
; Sequence 3, Application US/08714168
; Patent No. 5948636
;  GENERAL INFORMATION:
;    APPLICANT:  Hideharu MORI et al.
;    TITLE OF INVENTION:  CELL SURFACE POLYPEPTIDE GENE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,168
;      FILING DATE:  September 27, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1782 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Porphyromonas (Bacteroides) gingivalis
;      STRAIN:  OMZ409
;    FEATURE:
;      NAME/KEY:  -35 signal
;      LOCATION:  9..14
;      IDENTIFICATION METHOD:  S
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  -10 signal
;      LOCATION:  32..37
;      IDENTIFICATION METHOD:  S
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  60..1752
;      IDENTIFICATION METHOD:  S
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  Sig peptide
;      LOCATION:  60..105
;      IDENTIFICATION METHOD:  S
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  mat peptide
;      LOCATION:  208..1752
;      IDENTIFICATION METHOD:  E
;      OTHER INFORMATION:
US-08-714-168-3


  Query Match             3.3%;  Score 29.2;  DB 2;  Length 1782;
  Best Local Similarity   50.0%;  Pred. No. 20;
  Matches   73;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     201 tcccgaccccaaaacggtcaaagccgtttttgcggaaacggcaaaaggcggtttggaact 260
           | | | |  ||| ||||| ||||    | | | |   |  |  |  | |||   || |
Db     435 TGCTGTCGGCAAGACGGTTAAAGTATATGTGGTGCTCAATGACATCGCCGGCAAGGCAAA 494

Qy     261 tgcccccgcgtttttcagaaaaccggaagacatagaaacaatgttcaaagcggtacacgg 320
            ||||   |          ||  | | |||| | ||  | |  |||||||| ||||  ||
Db     495 AGCCCTTTTGGCAAATGTTAATGCAGCAGACTTTGATGCTAAATTCAAAGAGGTAATCGA 554

Qy     321 ctgggaacatgtgcagcaggctttgg 346
             |     |  |   |||||||| |
Db     555 ACTGTCTACTCAGGCTCAGGCTTTAG 580


Search completed: July  1, 2002, 02:30:47
Job time: 106372 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 08:42:47 ; Search time 27321.5 Seconds
(without alignments)
1279.881 Million cell updates/sec

Title: US-09-303-518D-465
Perfect score: 1671
Sequence: 1 ttgggcatttcccgcaaaat.....gaaagattacacacaaatga 1671

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1671	100.0	1671	6	A96432	A96432 Sequence 46
2	1671	100.0	311321	1	NMA3Z2491	AL162754 Neisseria
3	1248.2	74.7	1827	6	AX044009	AX044009 Sequence
4	1248.2	74.7	11796	1	AE002419	AE002419 Neisseria
5	1248.2	74.7	349980	6	AX044030	AX044030 Sequence
6	1231.2	73.7	1887	6	A96430	A96430 Sequence 46
7	1176.2	70.4	2028	6	AX236421	AX236421 Sequence
8	1176.2	70.4	2028	6	AX239731	AX239731 Sequence
9	1176.2	70.4	4425	6	AX236407	AX236407 Sequence
10	1176.2	70.4	4425	6	AX239715	AX239715 Sequence
11	1175.6	70.4	2019	6	AX236447	AX236447 Sequence
12	1175.6	70.4	2019	6	AX239733	AX239733 Sequence
13	1175.6	70.4	2256	6	AX236451	AX236451 Sequence
14	1175.6	70.4	2256	6	AX239737	AX239737 Sequence
15	1175.6	70.4	2421	6	AX236449	AX236449 Sequence
16	1175.6	70.4	2421	6	AX239735	AX239735 Sequence
17	1174.4	70.3	2304	6	AX236465	AX236465 Sequence
18	1174.4	70.3	2304	6	AX239751	AX239751 Sequence
19	1174.2	70.3	2256	6	AX236459	AX236459 Sequence
20	1174.2	70.3	2256	6	AX239745	AX239745 Sequence
21	1174.2	70.3	2421	6	AX236453	AX236453 Sequence
22	1174.2	70.3	2421	6	AX239739	AX239739 Sequence
23	653.4	39.1	683	6	A96426	A96426 Sequence 45
c 24	294.4	17.6	11700	1	AE002393	AE002393 Neisseria
c 25	294.4	17.6	349980	6	AX044029	AX044029 Sequence
26	286	17.1	195767	1	NMA7Z2491	AL162758 Neisseria
27	237.2	14.2	10057	1	AE002420	AE002420 Neisseria
28	121.8	7.3	1497	6	A96138	A96138 Sequence 17
29	121.8	7.3	1744	6	A61829	A61829 Sequence 9
30	121.8	7.3	3287	6	A61821	A61821 Sequence 1
31	121.8	7.3	3294	1	AF142582	AF142582 Neisseria
c 32	110.2	6.6	340806	1	NMA1Z2491	AL162752 Neisseria
33	107.2	6.4	1449	6	A96134	A96134 Sequence 16
34	107	6.4	1452	6	A96132	A96132 Sequence 16
35	107	6.4	10869	1	AE002559	AE002559 Neisseria
36	107	6.4	172325	6	AX044035	AX044035 Sequence
37	98	5.9	375	6	A96130	A96130 Sequence 16
38	51.8	3.1	125020	9	AF429315	AF429315 Homo sapi
39	45.4	2.7	10947	1	AE002514	AE002514 Neisseria
40	45.4	2.7	326301	1	NMA6Z2491	AL162757 Neisseria
41	45.4	2.7	349980	6	AX044033	AX044033 Sequence
c 42	42.2	2.5	3619	1	SCVALSFP	Y13070 S.coelicolo
43	42.2	2.5	42816	1	SCC88	AL139298 Streptomy
c 44	42.2	2.5	70423	2	AC102254	AC102254 Mus muscu
45	42.2	2.5	98980	2	AP004215	AP004215 Oryza sat
c 46	42.2	2.5	186323	2	AP004348	AP004348 Oryza sat
47	41.8	2.5	224448	2	PFMAL4P4	AL035477 Plasmodiu
48	40.8	2.4	2653	1	XXGDH	X16399 Gene for gl
49	40.8	2.4	172269	2	AC106508	AC106508 Rattus no
50	40.8	2.4	188489	2	AC016298	AC016298 Homo sapi
c 51	40.8	2.4	277050	1	AP000989	AP000989 Sulfolobu
52	40.6	2.4	605	8	AF377524	AF377524 Zea mays
53	40.6	2.4	605	8	AF377525	AF377525 Zea mays
54	40.6	2.4	605	8	AF377527	AF377527 Zea mays
55	40.6	2.4	605	8	AF377536	AF377536 Zea mays
56	40.6	2.4	605	8	AF377542	AF377542 Zea mays
c 57	40.4	2.4	7218	6	I66494	I66494 Sequence 14
c 58	40.2	2.4	171847	2	AL671007	AL671007 Mus muscu
c 59	40	2.4	18282	1	AF141932	AF141932 Rhizobium
c 60	40	2.4	122050	2	AC091512	AC091512 Leishmani
c 61	40	2.4	347660	1	AP002994	AP002994 Mesorhizo
62	39.8	2.4	390	6	AR135147	AR135147 Sequence
63	39.8	2.4	942	1	AF169562	AF169562 Mesorhizo
c 64	39.8	2.4	12026	1	AE005109	AE005109 Halobacte
c 65	39.6	2.4	3386	8	AB012641	AB012641 Nicotiana
66	39.6	2.4	15348	1	AE007163	AE007163 Mycobacte
67	39.6	2.4	47852	1	MTV023	AL022022 Mycobacte
c 68	39.6	2.4	110000	2	LMFLCHR18_00	AL499620 Leishmani
69	39.6	2.4	146519	2	AC074282	AC074282 Oryza sat
c 70	39.4	2.4	465	3	AF069170	AF069170 Dyscritul
71	39.4	2.4	162042	2	AC108206	AC108206 Homo sapi
72	39.2	2.3	968	1	ECOPROC	J01665 E.coli proC
73	39.2	2.3	10793	1	AE005758	AE005758 Caulobact

	74	39.2	2.3	15311	1	AE007164 AE007164 Mycobacte
	75	39.2	2.3	156573	2	AC073075 AC073075 Homo sapi
c	76	39.2	2.3	165339	3	CEY57A10A AL117195 Caenorhab
	77	39	2.3	605	8	AF377520 AF377520 Zea mays
	78	39	2.3	605	8	AF377521 AF377521 Zea mays
	79	39	2.3	605	8	AF377522 AF377522 Zea mays
	80	39	2.3	605	8	AF377523 AF377523 Zea mays
	81	39	2.3	605	8	AF377526 AF377526 Zea mays
	82	39	2.3	605	8	AF377529 AF377529 Zea mays
	83	39	2.3	605	8	AF377530 AF377530 Zea mays
	84	39	2.3	605	8	AF377531 AF377531 Zea mays
	85	39	2.3	605	8	AF377532 AF377532 Zea mays
	86	39	2.3	605	8	AF377533 AF377533 Zea mays
	87	39	2.3	605	8	AF377537 AF377537 Zea mays
	88	39	2.3	605	8	AF377538 AF377538 Zea mays
	89	39	2.3	605	8	AF377540 AF377540 Zea mays
	90	39	2.3	1696	8	ZMU14599 U14599 Zea mays Bz
c	91	39	2.3	5384	6	AX345110 AX345110 Sequence
	92	39	2.3	36224	1	SCD78 AL034355 Streptomy
	93	39	2.3	163786	2	AC097105 AC097105 Homo sapi
	94	38.8	2.3	10732	6	E32986 E32986 Gene encodi
	95	38.8	2.3	315000	1	RME603644 AL603644 Rhizobium
	96	38.6	2.3	882	6	AR169213 AR169213 Sequence
	97	38.6	2.3	882	6	BD006386 BD006386 Compounds
	98	38.6	2.3	882	6	BD006506 BD006506 Compounds
	99	38.6	2.3	1896	1	AF009829 AF009829 Mycobacte
	100	38.6	2.3	14707	1	AE006955 AE006955 Mycobacte

ALIGNMENTS

```
RESULT   1
A96432
LOCUS       A96432                   1671 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 465 from Patent WO9924578.
ACCESSION   A96432
VERSION     A96432.1  GI:6780108
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1671)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 465 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1671
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      484 a    440 c    402 g    345 t
ORIGIN
```

Query Match 100.0%; Score 1671; DB 6; Length 1671;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 60

Qy      61 catgcacacgcctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 120

Qy     121 catttcgaacccgacgggaaataccacctattcggcagcagggggaacttgccgagcgc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGAACTTGCCGAGCGC 180

Qy     181 agcggtcatatcggattgggaaacatacaaagccatcagttgggcaacctgttcatccag 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGCGGTCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCAACCTGTTCATCCAG 240

Qy     241 caggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaa 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 300

Qy     301 gtccattccccccttcgacaaccatgcctcacattccgattctgatgaagccggtagtccc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 360

Qy     361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 420

Qy     421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 480

Qy     481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 540

Qy     541 accggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagta 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ACCGGACAACGGCTTGTCGACCGTTTCCACAATACCGGTAGTATGCTGACGCAAGGAGTA 600

Qy     601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 660

Qy     661 gccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCCGAAGCTTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAA 720

Qy     721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatg 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTTATG 780

Qy     781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 840

Qy     841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900

Qy     901 aatgccgcacaaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaa 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTACGGCAGTCATCCCCGTCAAA 960

Qy     961 gggattggagctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAG 1020

Qy    1021 cggtcgcagatgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGGTCGCAGATGGGCGAGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 1080

Qy    1081 gccgatgcggcatacgccaaataccgtccccttaccattcccgaaatatccgttcaaac 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 1140

Qy    1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacgga 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGA 1200

Qy    1201 aagaatgtgaaactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 AAGAATGTGAAACTGGCAAACAAACGCCACCCGAAGACCAAAGTGCCGTTTGACGGTAAA 1260

Qy    1261 gggtttccgaattttgaaaaagacgtaaaatacgatacgagaattaataccgctgtacca 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1261 GGGTTTCCGAATTTTGAAAAAGACGTAAAATACGATACGAGAATTAATACCGCTGTACCA 1320

Qy   1321 caagtgaatcctatagatgaacccgtctttaatcctaaaggttctgtcggatcggctcat 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CAAGTGAATCCTATAGATGAACCCGTCTTTAATCCTAAAGGTTCTGTCGGATCGGCTCAT 1380

Qy   1381 tcttggtctataactgccagaattcaatacgcaaaattaccaaggcaaggtagaatcaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TCTTGGTCTATAACTGCCAGAATTCAATACGCAAAATTACCAAGGCAAGGTAGAATCAGA 1440

Qy   1441 tatatcccacctaaaaattactctccttcagcaccgctaccaaaaggacctaataatgga 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TATATCCCACCTAAAAATTACTCTCCTTCAGCACCGCTACCAAAAGGACCTAATAATGGA 1500

Qy   1501 tatttggataaatttggtaatgaatggactaaaggtccatcaagaactaaaggtcaagaa 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TATTTGGATAAATTTGGTAATGAATGGACTAAAGGTCCATCAAGAACTAAAGGTCAAGAA 1560

Qy   1561 tttgaatgggatgttcaattgtctaaaacaggaagagagcaacttggatgggctagtagg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTTGAATGGGATGTTCAATTGTCTAAAACAGGAAGAGAGCAACTTGGATGGGCTAGTAGG 1620

Qy   1621 gatggtaagcatttaaatatatcaattgatggaaagattacacacaaatga 1671
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GATGGTAAGCATTTAAATATATCAATTGATGGAAAGATTACACACAAATGA 1671


RESULT   2
NMA3Z2491
LOCUS       NMA3Z2491              311321 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 3/7.
ACCESSION   AL162754 AL157959
VERSION     AL162754.2  GI:7379424
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 311321)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 311321)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .311321
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     RBS             115. .120
     gene            129. .587
                     /gene="NMA0698"
     CDS             129. .587
                     /gene="NMA0698"
                     /note="NMA0698, possible periplasmic protein, len: 152 aa;
                     unknown, contains a probable N-terminal signal sequence
                     and lies within a region of unusually low GC content"
                     /codon_start=1
                     /transl_table=11
                     /product="putative periplasmic protein"
                     /protein_id="CAB83985.1"
                     /db_xref="GI:7379425"
                     /db_xref="SPTREMBL:Q9JQQ2"
                     /translation="MLTKLKILLFLFLFVFVLAINLLFFFFSSDIESFGNYQFEYVYD
                     KGWPANYILVMKDGNEGNFDKIISGLVLEYYKEDDNIYFSYIDGQGFASDSCYYKPEI
                     LYGKIILNKNHIININSMEKNNFLSEDKIMKGTRNWLADPKNKCNIQTLD"
     gene            742. .1077
                     /gene="NMA0699"
     CDS             742. .1077
                     /gene="NMA0699"
                     /note="NMA0699, pseudogene, probable ABC transporter
                     protein, len: 336 bp; similar to C-termini of many ABC
                     transporters e.g. TR:Q46973 (EMBL:U47048), mtfB,
                     Escherichia coli microcin transport protein (707 aa),
                     fasta scores; E(): 2.6e-19, 53.2% identity in 111 aa
                     overlap. Contains PS00211 ABC transporters family
                     signature. NMA0686 may be the remainder of this
                     pseudogene"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="putative ABC transporter protein (pseudogene)"
     misc_feature    868. .912
                     /gene="NMA0699"
                     /note="PS00211 ABC transporters family signature"
     gene            complement(1101. .2327)
                     /gene="NMA0700"
     CDS             complement(1101. .2327)
                     /gene="NMA0700"
                     /note="NMA0700, possible ribonuclease BN, len: 408 aa;
                     similar to SW:RBN_ECOLI (EMBL:L19201), rbn, Escherichia
                     coli ribonuclease BN (EC 3.1.-.-), fasta scores; E():
                     2.3e-28, 34.3% identity in 268 aa overlap. Longer than rbn
                     at the C-terminus. Also similar to TR:O85449
                     (EMBL:AF067083) Vitreoscilla sp. hypothetical protein (376
                     aa), fasta scores; E(): 0, 45.0% identity in 333 aa
                     overlap. Contains a region similar to NMA0677, fasta
                     scores; E(): 4.3e-08, 64.4% identity in 45 aa overlap.
                     Contains hydrophobic, possible membrane-spanning regions,
                     as does rbn"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribonuclease BN"
                     /protein_id="CAB83987.1"
                     /db_xref="GI:7379426"
                     /db_xref="SPTREMBL:Q9JVT9"
                     /translation="MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLL
                     ALVPVLTVMVAVASIFPVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTA
                     IGSVMLVVTSLMLIRTIDNTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFM
                     VGSVQDAALASGAPQWSGALRTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATA
                     FCLETARSLFTWYMGNFDGYRSIYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQ
                     GEAFRRGFDSRGRFDDVLKILLLLDAAQKEGKALPVQEFRRHINMGYDELGELLEKLA
                     RHGYIYSGRQGWVLKTGADSIELNELFKLFVYRPLPVERDHVNQAVDAVMMPCLQTLN
                     MTLAEFDAQAKKQQQS"
     misc_feature    1376. .1385
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(2417. .2426)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2444. .2453
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            2486. .2743
                     /gene="NMA0701"
     CDS             2486. .2743
                     /gene="NMA0701"
                     /note="NMA0701, possible pseudogene, len: 258 bp; shows
                     weak similarity to part of SW:WRBA_ECOLI (EMBL:M99166),
```

wrbA, Escherichia coli Trp repressor binding protein (197 aa), fasta scores; E(): 0.064, 29.3% identity in 82 aa overlap. Also similar to part of TR:O85450 (EMBL:AF067083) Vitreoscilla sp. Trp repressor binding protein (fragment) (124 aa), fasta scores; E(): 6.6e-14, 56.6% identity in 83 aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="hypothetical protein NMA0701 (pseudogene)"

misc_feature complement(2755. .2764)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

misc_feature 2778. .2787
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

RBS 2861. .2864

gene 2871. .3530
/gene="NMA0702"

CDS 2871. .3530
/gene="NMA0702"
/note="NMA0702, len: 219 aa; similar to TR:Q52925 (EMBL:Z50189), exsB, Rhizobium meliloti putative regulator of succinoglycan biosynthesis (not a transcriptional regulator) (234 aa), fasta scores; E(): 3.1e-15, 32.7% identity in 205 aa overlap. Also similar to many hbacterial ypothetical proteins e.g. SW:YBAX_HAEIN (EMBL:U32798), HI1191, Haemophilus influenzae hypothetical protein (196 aa), fasta scores; E(): 0, 79.2% identity in 173 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0702"
/protein_id="CAB83989.1"
/db_xref="GI:7379427"
/db_xref="SPTREMBL:Q9JVT8"
/translation="MSNQKALVIFSGGQDSTTCLIQAIQTYGRENVQAITFQYGQRHA VELERARWIAQDLGVKQTVLDLSLMRQITHNALMDDTAAIETAENGVPNTFVDGRNAL FLLYAAIYAKGQGIRHIIAGVCETDFSGYPDCRDVFVKSMNVTLNLAMDYDFQIHTPL MYLTKAQTWALADEMGALDYIREQTHTCYNGIVGGCRECPSCILRERGLAEYLESKKA V"

misc_feature 3522. .3531
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

RBS 3547. .3550

gene 3558. .4076
/gene="NMA0703"

CDS 3558. .4076
/gene="NMA0703"
/note="NMA0703, len: 172 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0703"
/protein_id="CAB83990.1"
/db_xref="GI:7379428"
/db_xref="SPTREMBL:Q9JVT7"
/translation="MPKLHMFYLGGNAGRSNIEVHDIQFAVCDDYREAVPALKAAWFG DADKIHIDGWQVVEWADGYDIAVSETPKTKMPSENAPRLYFANVGGYRAGQLAEAHAF GLFAAATPAEAKQKALQTLLTDSYVQQHKDNLKDVDNLLALDRIGNFHIRLTPNPHGK PAEIGFQGYLPI"

misc_feature 3782. .3791
/gene="NMA0703"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

gene 4083. .4505
/gene="NMA0704"

misc_feature 4083. .4502
/gene="NMA0704"
/note="Pfam match to entry PF01242 PTPS, 6-pyruvoyl tetrahydropterin synthase, score 10.80, E-value 1.7e-06"

CDS 4083. .4505
/gene="NMA0704"
/note="NMA0704, len: 140 aa; similar to many hypothetical proteins e.g. SW:YB90_HAEIN (EMBL:U32798), HI1190, Haemophilus influenzae hypothetical protein (141 aa), fasta scores; E(): 5.1e-32, 58.0% identity in 138 aa overlap. Shows very weak similarity to eukaryotic 6-pyruvoyl-tetrahydropterin synthases e.g. SW:PTPS_RAT (EMBL:M77850), pts, Rattus norvegicus 6-pyruvoyl tetrahydrobiopterin synthase precursor (EC 4.6.1.10) (144 aa), fasta scores; E(): 0.0062, 27.3% identity in 154 aa overlap. Contains Pfam match to entry PF01242 PTPS, 6-pyruvoyl tetrahydropterin synthase"

Query Match 100.0%; Score 1671; DB 1; Length 311321;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142661 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 142720

Qy      61 catgcacacgcctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142721 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 142780

Qy     121 catttcgaacccgacgggaaataccacctattcggcagcagggggggaacttgccgagcgc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142781 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGC 142840

Qy     181 agcggtcatatcggattgggaaacatacaaagccatcagttgggcaacctgttcatccag 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142841 AGCGGTCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCAACCTGTTCATCCAG 142900

Qy     241 caggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaa 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142901 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 142960

Qy     301 gtccattccccccttcgacaaccatgcctcacattccgattctgatgaagccggtagtccc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142961 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 143020

Qy     361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143021 GTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 143080

Qy     421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143081 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 143140

Qy     481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143141 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 143200

Qy     541 accggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagta 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143201 ACCGGACAACGGCTTGTCGACCGTTTCCACAATACCGGTAGTATGCTGACGCAAGGAGTA 143260

Qy     601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143261 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 143320

Qy     661 gccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143321 GCCGAAGCTTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAA 143380

Qy     721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatg 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143381 ATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTTATG 143440

Qy     781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143441 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 143500

Qy     841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 143501 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 143560

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143561 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTACGGCAGTCATCCCCGTCAAA 143620

Qy    961 gggattggagctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143621 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAG 143680

Qy   1021 cggtcgcagatgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143681 CGGTCGCAGATGGGCGAGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 143740

Qy   1081 gccgatgcggcatacgccaaatacccgtccccttaccattcccgaaatatccgttcaaac 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143741 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 143800

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacgga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143801 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGA 143860

Qy   1201 aagaatgtgaaactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143861 AAGAATGTGAAACTGGCAAACAAACGCCACCCGAAGACCAAAGTGCCGTTTGACGGTAAA 143920

Qy   1261 gggtttccgaattttgaaaaagacgtaaaatacgatacgagaattaataccgctgtacca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143921 GGGTTTCCGAATTTTGAAAAAGACGTAAAATACGATACGAGAATTAATACCGCTGTACCA 143980

Qy   1321 caagtgaatcctatagatgaacccgtctttaatcctaaaggttctgtcggatcggctcat 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143981 CAAGTGAATCCTATAGATGAACCCGTCTTTAATCCTAAAGGTTCTGTCGGATCGGCTCAT 144040

Qy   1381 tcttggtctataactgccagaattcaatacgcaaaattaccaaggcaaggtagaatcaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 144041 TCTTGGTCTATAACTGCCAGAATTCAATACGCAAAATTACCAAGGCAAGGTAGAATCAGA 144100

Qy   1441 tatatcccacctaaaaattactctccttcagcaccgctaccaaaaggacctaataatgga 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 144101 TATATCCCACCTAAAAATTACTCTCCTTCAGCACCGCTACCAAAAGGACCTAATAATGGA 144160

Qy   1501 tatttggataaatttggtaatgaatggactaaaggtccatcaagaactaaaggtcaagaa 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 144161 TATTTGGATAAATTTGGTAATGAATGGACTAAAGGTCCATCAAGAACTAAAGGTCAAGAA 144220

Qy   1561 tttgaatgggatgttcaattgtctaaaacaggaagagagcaacttggatgggctagtagg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 144221 TTTGAATGGGATGTTCAATTGTCTAAAACAGGAAGAGAGCAACTTGGATGGGCTAGTAGG 144280

Qy   1621 gatggtaagcatttaaatatatcaattgatggaaagattacacacaaatga 1671
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 144281 GATGGTAAGCATTTAAATATATCAATTGATGGAAAGATTACACACAAATGA 144331


RESULT   3
AX044009
LOCUS       AX044009               1827 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 88 from Patent WO0066791.
ACCESSION   AX044009
VERSION     AX044009.1  GI:11342894
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1827)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 88 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES             Location/Qualifiers
     source          1. .1827
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
BASE COUNT      539 a    446 c    454 g    387 t      1 others
ORIGIN


  Query Match             74.7%;  Score 1248.2;  DB 6;  Length 1827;
  Best Local Similarity   97.1%;  Pred. No. 0;
  Matches 1271;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 60

Qy     61 catgcacacgcctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 120

Qy    121 catttcgaacccgacgggaaataccacctattcggcagcagggggggaacttgccgagcgc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGC 180

Qy    181 agcggtcatatcggattgggaaacatacaaagccatcagttgggcaacctgttcatccag 240
          ||||| |||||||||||||||| |||||||||||||||||||||||||| | || ||
Db    181 AGCGGCCATATCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAA 240

Qy    241 caggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 300

Qy    301 gtccattccccttcgacaaccatgcctcacattccgattctgatgaagccggtagtccc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 360

Qy    361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTTGACGGATTTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 420

Qy    421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 480

Qy    481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 540

Qy    541 accggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagta 600
          |||||||||||||| |||||||||||||||| |||||||||||||||||||||||||||
Db    541 ACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTA 600

Qy    601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 660

Qy    661 gccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
          |||||||| |||||||||||||||||||| |||||||||||||||||||||||||||||
Db    661 GCCGAAGCCTTCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAA 720

Qy    721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatg 780
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||| |||
Db    721 ATTGTCGGCGCAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATG 780

Qy    781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 840

Qy    841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaa 960
```

```
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||| |||||
Db    901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAA 960

Qy    961 gggattggagctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    961 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAG 1020

Qy   1021 cggtcgcagatgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 1080

Qy   1081 gccgatgcggcatacgccaaatacccgtccccttaccattcccgaaatatccgttcaaac 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 1140

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacgga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 1200

Qy   1201 aagaatgtgaaactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaa 1260
          || ||||| ||||||||| || |||||||||||||||    || ||||||||||||||||
Db   1201 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAA 1260

Qy   1261 gggtttccgaattttgaaaaagacgtaaaatacgatacgagaattaata 1309
          |||||||||||||||| ||  |||| ||||| |||||||  |  |||
Db   1261 GGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATA 1309
```

```
RESULT   4
AE002419
LOCUS       AE002419               11796 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 61 of 206 of
            the complete genome.
ACCESSION   AE002419 AE002098
VERSION     AE002419.1  GI:7225863
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 11796)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 11796)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11796
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            complement(174. .1139)
                     /gene="NMB0635"
     CDS             complement(174. .1139)
                     /gene="NMB0635"
                     /note="similar to SP:P37247 PID:1197006 percent identity:
                     61.94; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase, IS30 family"
                     /protein_id="AAF41058.1"
                     /db_xref="GI:7225864"
                     /translation="MSYTQLTQGERYHIQYLSRHCTVTEIAKQLNRHKSTISREIRRH
                     RTQGQQYSAEKAQRQSRTIKQRKRQPYKLDSQLIQHIDPLIRRKLSPEQVCAYLCKHH
                     QITLHHSTIYRYLRQDKSNGSTLWQHLRICSKPYRKRYGSTWTRGKVPNRVGIENRPA
                     IVDQKSRIGDWEADTIVGKGQKSALLTLVERVTRYTIICKLDSLKAEDTARAAVRALK
                     AHKDRVHTITMDNGKEFYQHTKITKALKAETYFCRPYHSWEKGLNENTNGLIRQYFPK
                     QTDFRNISDREIRRVQDELNHRPRKTLGYETPSVLFLNLFQPLIH"
     gene            complement(1303. .1686)
                     /gene="NMB0636"
     CDS             complement(1303. .1686)
                     /gene="NMB0636"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41059.1"
                     /db_xref="GI:7225865"
                     /translation="MTDTDTQADRFEQMMWQAVDKLFEQHDGKLESMDGREQELVLIW
                     RTEADIGNGSILQFVCNWGFPAAEKTCSVLKKIGAVHSAMLIHRAADALDKEIRRLQS
                     EGKNLKEMWDITSRQQNRLTAEQSG"
     gene            complement(1744. .3120)
                     /gene="NMB0637"
     CDS             complement(1744. .3120)
                     /gene="NMB0637"
                     /note="similar to PID:1652362 percent identity: 67.97;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="argininosuccinate lyase"
                     /protein_id="AAF41060.1"
                     /db_xref="GI:7225866"
                     /translation="MHDKTWSGRFNEPVSELVKQYTASIGFDRRLAEWDIQGSLAHAQ
                     MLKETGVLDEGDLADIRRGMAEILEEIRSGKIEWSSDLEDVHMNIERRLTDKIGDAGK
                     RLHTGRSRNDQVATDIRLWLRDQITVIQSLIQSLQTALLDLAEQNAETVMPGFTHLQV
                     AQPVSFGHHMLAYVEMLGRDNERMADCRCRVNRMPLGAAALAGTTYPIQREITAELLG
                     FEQICQNSLDAVSDRDFAIEFTAAASLVMVHLSRLSEELILWMSPRFGFIDIADRFCT
                     GSSIMPQKKNPDVPELVRGKSGRVIGHLIGLITLMKSQPLAYNKDNQEDKEPLFDTAD
                     TLIDTLRIYADMMRGVTVKPDNMRAAVMQGFATATDLADYLVKKGMPFRDAHEVVAQA
                     VRHADQAGVDLSELPLEVLQGFSDLIADDVYGVLTPEGSLNARNHLGGTAPEQVRFQV
                     KRWREMLA"
     gene            complement(3139. .4008)
                     /gene="NMB0638"
     CDS             complement(3139. .4008)
                     /gene="NMB0638"
                     /note="similar to PID:1628575 percent identity: 74.13;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP--glucose-1-phosphate uridylyltransferase"
                     /protein_id="AAF41061.1"
                     /db_xref="GI:7225867"
                     /translation="MKPIRKAVFPVAGMGTRFLPATKASPKEMLPIVDKPLIQYAVEE
                     AVEAGCTEMVFVTGRNKRSIEDHFDKAYELETELEMRHKDKLLEHVRNILPPNITCLY
                     IRQAEALGLGHAVLCARAAIGDEPFAVILADDLIDAPKGALKQMVEVYGRSGNSILGV
                     ETVEASQTGSYGIVETEQLKQFQRITGIVEKPKPEDAPSNLAVVGRYILTPRIFDLLT
                     NLPRGAGNEIQLTDGIAKLLDHEFVLAHPFEGTRYDCGSKLGYLEATVAYGLKHPETG
                     EPFRRLLEKYRTE"
     gene            complement(4036. .4635)
                     /gene="NMB0639"
     CDS             complement(4036. .4635)
                     /gene="NMB0639"
```

```
                /note="conserved hypothetical protein; identified by
                Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF41062.1"
                /db_xref="GI:7225868"
                /translation="MSEKPEKIVLASGNAGKLEEFGNLFKPYSITVLPQSAFGIPECP
                EPYPTFVENALAKARHAAKYSGLPALADDSGICAAALNGAPGIHSARYAGDNPKSDTA
                NNLKLAAELVGKADKSCCYVCVLVFVRHKDDPRPIIAEGVWHGQWNDTPLGQNGFGYD
                PYFYLPEHGKTAAELDTEVKNRESHRAQALAELLRKLAL"
gene            complement(4628. .4861)
                /gene="NMB0640"
CDS             complement(4628. .4861)
                /gene="NMB0640"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF41063.1"
                /db_xref="GI:7225869"
                /translation="MNRRKIYLLSVALFTLAFMLLVLLGAYLLTVGSKAFAVASFLFA
                FGALFGQIGSLALYLRHKSLRAAQSATKENRYV"
gene            complement(4969. .5502)
                /gene="NMB0641"
CDS             complement(4969. .5502)
                /gene="NMB0641"
                /note="similar to GB:L42023 SP:P44529 PID:1003152
                PID:1222038 PID:1204383 percent identity: 92.61;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="inorganic pyrophosphatase"
                /protein_id="AAF41064.1"
                /db_xref="GI:7225870"
                /translation="MADFNQILTPGDVDGGIINVVNEIPAGSNHKIEWNRKLAAFQLD
                RVEPAIFAKPTNYGFIPQTLDEDGDELDVLLVTEQPLATGVFLEARVIGVMKFVDDGE
                VDDKIVCVPADDRNNGNAYKTLSDLPQQLIKQIEFHFNHYKDLKKAGTTKVESWGDAE
                EAKKVIKESIERWNKQA"
gene            complement(5603. .6061)
                /gene="NMB0642"
CDS             complement(5603. .6061)
                /gene="NMB0642"
                /note="similar to GB:D10165 SP:P24236 PID:581152
                PID:669113 PID:912430 percent identity: 58.45; identified
                by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="dATP pyrophosphohydrolase"
                /protein_id="AAF41065.1"
                /db_xref="GI:7225871"
                /translation="MPMAKPLKYPVSALVVLYSGDGGILLIERTHPEGFWQSVTGSLE
                PGETVAQTARREVWEETGILLADGQLQDWHDSTVYEIYHHWRHRYPKGVFENREHLFS
                AEIPRDTPIALQPEEHVSYGWFDMEEAAEKVFSPSNRRAILELGRFLGKR"
tRNA            6272. .6348
                /gene="tRNA-Pro-3"
                /product="tRNA-Pro"
gene            6272. .6348
                /gene="tRNA-Pro-3"
gene            6506. .8332
                /gene="NMB0643"
CDS             6506. .8332
                /gene="NMB0643"
                /note="MafB-related protein; identified by Oxford;
                putative"
                /codon_start=1
                /transl_table=11
                /product="MafB-related protein"
                /protein_id="AAF62313.1"
                /db_xref="GI:7413435"
                /translation="MGISRKISLILSILAVCLPMHAHASDLANDSFIRQVLDRQHFEP
                DGKYHLFGSRGELAERSGHIGLGKIQSHQLGNLMIQQAAIKGNIGYIVRFSDHGHEVH
                SPFDNHASHSDSDEAGSPVDGFSLYRIHWDGYEHHPADGYDGPQGGGYPAPKGARDIY
                SYDIKGVAQNIRLNLTDNRSTGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDRSG
                NAAEAFNGTADIVKNIIGAAGEIVGAGDAVQGISEGSNIAVMHGLGLLSTENKMARIN
                DLADMAQLKDYAAAAIRDWAVQNPNAAQGIEAVSNIFMAAIPIKGIGAVRGKYGLGGI
                TAHPIKRSQMGAIALPKGKSAVSDNFADAAYAKYPSPYHSRNIRSNLEQRYGKENITS
                STVPPSNGKNVKLADQRHPKTGVPFDGKGFPNFEKHVKYDTKLDIQELSGGGIPKAKP
                VFDAKPRWEVDRKLNKLTTREQVEKNVQEIRNGNINSNFSQHAQLEREINKLKSADEI
                NFADGMGKFTDSMNDKAFSRLVKSVKENGFTNPVVEYVEINGKAYIVRGNNRVFAAEY
                LGRIHELKFKKVDFPVPNTSWKNPTDVLNESGNVKRPRYRSK"
gene            8335. .8799
                /gene="NMB0644"
```

Query Match 74.7%; Score 1248.2; DB 1; Length 11796;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1271; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```
Qy       1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6506 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 6565

Qy      61 catgcacacgcctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6566 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 6625

Qy     121 catttcgaacccgacgggaaataccacctattcggcagcaggggggaacttgccgagcgc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6626 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGC 6685

Qy     181 agcggtcatatcggattgggaaacatacaaagccatcagttgggcaacctgttcatccag 240
           ||||| |||||||||||||||| |||||||||||||||||||||||||| | || ||
Db    6686 AGCGGCCATATCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAA 6745

Qy     241 caggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaa 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6746 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 6805

Qy     301 gtccattccccccttcgacaaccatgcctcacattccgattctgatgaagccggtagtccc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6806 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 6865

Qy     361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    6866 GTTGACGGATTTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 6925

Qy     421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6926 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 6985

Qy     481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6986 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 7045

Qy     541 accggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagta 600
           |||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||
Db    7046 ACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTA 7105

Qy     601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7106 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 7165

Qy     661 gccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
           |||||||| |||||||||||||||||||||| |||||||||||||||||||||||||||
Db    7166 GCCGAAGCCTTCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAA 7225

Qy     721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatg 780
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||| |||
Db    7226 ATTGTCGGCGCAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATG 7285

Qy     781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7286 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 7345

Qy     841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
```

```
Db   7346 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 7405

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaa 960
Db   7406 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAA 7465

Qy    961 gggattggagctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
Db   7466 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAG 7525

Qy   1021 cggtcgcagatgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
Db   7526 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 7585

Qy   1081 gccgatgcggcatacgccaaataccgtcccttaccattcccgaaatatccgttcaaac 1140
Db   7586 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 7645

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacgga 1200
Db   7646 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 7705

Qy   1201 aagaatgtgaaactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaa 1260
Db   7706 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAA 7765

Qy   1261 gggtttccgaattttgaaaaagacgtaaaatacgatacgagaattaata 1309
Db   7766 GGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATA 7814


RESULT   5
AX044030
LOCUS       AX044030               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 109 from Patent WO0066791.
ACCESSION   AX044030
VERSION     AX044030.1  GI:11342914
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G.; Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 109 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    82523 a  82940 c  96712 g  87805 t
ORIGIN


  Query Match           74.7%;  Score 1248.2;  DB 6;  Length 349980;
  Best Local Similarity   97.1%;  Pred. No. 0;
  Matches 1271;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
Db  75614 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 75673

Qy     61 catgcacacgcctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcag 120
Db  75674 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 75733

Qy    121 catttcgaacccgacgggaaataccacctattcggcagcagggggaacttgccgagcgc 180
Db  75734 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGAACTTGCCGAGCGC 75793

Qy    181 agcggtcatatcggattgggaaacatacaaagccatcagttgggcaacctgttcatccag 240
Db  75794 AGCGGCCATATCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAA 75853

Qy    241 caggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaa 300
Db  75854 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 75913

Qy    301 gtccattccccccttcgacaaccatgcctcacattccgattctgatgaagccggtagtccc 360
Db  75914 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 75973

Qy    361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
Db  75974 GTTGACGGATTTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 76033

Qy    421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
Db  76034 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 76093

Qy    481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
Db  76094 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 76153

Qy    541 accggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagta 600
Db  76154 ACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTA 76213

Qy    601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
Db  76214 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 76273

Qy    661 gccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
Db  76274 GCCGAAGCCTTCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAA 76333

Qy    721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatg 780
Db  76334 ATTGTCGGCGCAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATG 76393

Qy    781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
Db  76394 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 76453

Qy    841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
Db  76454 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 76513

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaa 960
Db  76514 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAA 76573

Qy    961 gggattggagctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
Db  76574 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAG 76633

Qy   1021 cggtcgcagatgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
Db  76634 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 76693

Qy   1081 gccgatgcggcatacgccaaataccgtcccttaccattcccgaaatatccgttcaaac 1140
Db  76694 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 76753

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacgga 1200
```

```
Db  76754 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 76813

Qy   1201 aagaatgtgaaactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaa 1260
Db  76814 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAA 76873

Qy   1261 gggtttccgaattttgaaaaagacgtaaaatacgatacgagaattaata 1309
Db  76874 GGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATA 76922


RESULT   6
A96430
LOCUS       A96430                 1887 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 463 from Patent WO9924578.
ACCESSION   A96430
VERSION     A96430.1  GI:6780107
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1887)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 463 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1887
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      554 a    468 c    493 g    371 t      1 others
ORIGIN


  Query Match           73.7%;  Score 1231.2;  DB 6;  Length 1887;
  Best Local Similarity  96.3%;  Pred. No. 6.1e-311;
  Matches 1260;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy      1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
Db      1 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 60

Qy     61 catgcacacgcctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcag 120
Db     61 CATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAG 120

Qy    121 catttcgaacccgacgggaaataccacctattcggcagcagggggaacttgccgagcgc 180
Db    121 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGC 180

Qy    181 agcggtcatatcggattgggaaacatacaaagccatcagttgggcaacctgttcatccag 240
Db    181 AACGGCCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAA 240

Qy    241 caggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaa 300
Db    241 CAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAA 300

Qy    301 gtccattccccttcgacaaccatgcctcacattccgattctgatgaagccggtagtccc 360
Db    301 TTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCC 360

Qy    361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
Db    361 GTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 420

Qy    421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
Db    421 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 480

Qy    481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
Db    481 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 540

Qy    541 accggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagta 600
Db    541 ACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTA 600

Qy    601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
Db    601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 660

Qy    661 gccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
Db    661 GCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAA 720

Qy    721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatg 780
Db    721 ATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATG 780

Qy    781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
Db    781 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 840

Qy    841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
Db    841 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaa 960
Db    901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAA 960

Qy    961 gggattggagctgttccggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
Db    961 GGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAG 1020

Qy   1021 cggtcgcagatgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
Db   1021 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 1080

Qy   1081 gccgatgcggcatacgccaaatacccgtcccccttaccattcccgaaatatccgttcaaac 1140
Db   1081 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 1140

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacgga 1200
Db   1141 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 1200

Qy   1201 aagaatgtgaaactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaa 1260
Db   1201 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAA 1260

Qy   1261 gggtttccgaattttgaaaaagacgtaaaatacgatacgagaattaata 1309
Db   1261 GGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATA 1309


RESULT   7
AX236421
LOCUS       AX236421               2028 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 114 from Patent WO0164922.
ACCESSION   AX236421
VERSION     AX236421.1  GI:15796033
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2028)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 114 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
```

```
     source          1. .2028
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="deltaG741-ORF46.1"
BASE COUNT       557 a    577 c    531 g    363 t
ORIGIN


  Query Match             70.4%;  Score 1176.2;  DB 6;  Length 2028;
  Best Local Similarity   97.3%;  Pred. No. 1.5e-296;
  Matches 1196;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      71 cctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaac 130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     773 CCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 832

Qy     131 ccgacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcata 190
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     833 CCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATA 892

Qy     191 tcggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggcca 250
           |||||||||||| |||||||||||||||||||||||||| | || || ||||||||||
Db     893 TCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCA 952

Qy     251 ttaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccc 310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     953 TTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCC 1012

Qy     311 ccttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggat 370
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1013 CCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGAT 1072

Qy     371 tcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacg 430
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1073 TTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACG 1132

Qy     431 ggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgaca 490
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1133 GGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACA 1192

Qy     491 taaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaac 550
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1193 TAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 1252

Qy     551 ggcttgtcgaccgtttccacaatacggtagtatgctgacgcaaggagtaggcgacggat 610
           |||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    1253 GGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGAT 1312

Qy     611 tcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctt 670
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    1313 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCT 1372

Qy     671 tcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcg 730
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    1373 TCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 1432

Qy     731 caggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgg 790
           |||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||||
Db    1433 CAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 1492

Qy     791 gtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaac 850
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1493 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 1552

Qy     851 tcaaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcac 910
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1553 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 1612

Qy     911 aaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggag 970
           |||||||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||
Db    1613 AAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 1672

Qy     971 ctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcaga 1030
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1673 CTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGA 1732

Qy    1031 tgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcgg 1090
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1733 TGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG 1792

Qy    1091 catacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagc 1150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1793 CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 1852

Qy    1151 gttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtga 1210
           |||||||||||||||||||||||||||||||||||||||||||||||| || ||||| |
Db    1853 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCA 1912

Qy    1211 aactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccga 1270
           |||||||| || |||||||||||||||     || ||||||||||||||||||||||||||
Db    1913 AACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 1972

Qy    1271 attttgaaaaagacgtaaaatacgatacg 1299
           ||||||| ||  |||| ||||| ||||||
Db    1973 ATTTTGAGAAGCACGTGAAATATGATACG 2001


RESULT   8
AX239731
LOCUS       AX239731               2028 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 33 from Patent WO0164920.
ACCESSION   AX239731
VERSION     AX239731.1  GI:15797376
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2028)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 33 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2028
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="deltaG741-ORF46.1"
BASE COUNT       557 a    577 c    531 g    363 t
ORIGIN


  Query Match             70.4%;  Score 1176.2;  DB 6;  Length 2028;
  Best Local Similarity   97.3%;  Pred. No. 1.5e-296;
  Matches 1196;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      71 cctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaac 130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     773 CCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 832

Qy     131 ccgacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcata 190
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     833 CCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATA 892

Qy     191 tcggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggcca 250
           |||||||||||| |||||||||||||||||||||||||| | || || ||||||||||
Db     893 TCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCA 952

Qy     251 ttaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccc 310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     953 TTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCC 1012

Qy     311 ccttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggat 370
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1013 CCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGAT 1072
```

```
Qy    371 tcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacg 430
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1073 TTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACG 1132

Qy    431 ggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgaca 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 GGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACA 1192

Qy    491 taaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaac 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 TAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 1252

Qy    551 ggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggat 610
          |||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1253 GGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGAT 1312

Qy    611 tcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagcctt 670
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   1313 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCT 1372

Qy    671 tcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcg 730
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1373 TCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 1432

Qy    731 caggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgg 790
          |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db   1433 CAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 1492

Qy    791 gtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaac 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1493 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 1552

Qy    851 tcaaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcac 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1553 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 1612

Qy    911 aaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggag 970
          |||||||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||
Db   1613 AAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 1672

Qy    971 ctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcaga 1030
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1673 CTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGA 1732

Qy   1031 tgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcgg 1090
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1733 TGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG 1792

Qy   1091 catacgccaaataccgtccccttaccattcccgaaatatccgttcaaacttggagcagc 1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1793 CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 1852

Qy   1151 gttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtga 1210
          |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||||||
Db   1853 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCA 1912

Qy   1211 aactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccga 1270
          |||||||| || ||||||||||||||||    || |||||||||||||||||||||||||||
Db   1913 AACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 1972

Qy   1271 attttgaaaaagacgtaaaatacgatacg 1299
          ||||||| ||  |||| ||||| ||||||
Db   1973 ATTTTGAGAAGCACGTGAAATATGATACG 2001
```

```
RESULT   9
AX236407
LOCUS       AX236407               4425 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 100 from Patent WO0164922.
ACCESSION   AX236407
VERSION     AX236407.1  GI:15796026
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 4425)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 100 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .4425
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="deltaG983-ORF46.1"
BASE COUNT     1165 a   1259 c   1219 g    782 t
ORIGIN
```

Query Match 70.4%; Score 1176.2; DB 6; Length 4425;
Best Local Similarity 97.3%; Pred. No. 1.8e-296;
Matches 1196; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy     71 cctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaac 130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3170 CCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 3229

Qy    131 ccgacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcata 190
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   3230 CCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATA 3289

Qy    191 tcggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggcca 250
          |||||||||||||| |||||||||||||||||||||||||| | || || ||||||||||
Db   3290 TCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCA 3349

Qy    251 ttaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccc 310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3350 TTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCC 3409

Qy    311 ccttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggat 370
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3410 CCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGAT 3469

Qy    371 tcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacg 430
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3470 TTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACG 3529

Qy    431 ggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgaca 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3530 GGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACA 3589

Qy    491 taaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaac 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3590 TAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 3649

Qy    551 ggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggat 610
          |||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3650 GGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGAT 3709

Qy    611 tcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagcctt 670
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   3710 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCT 3769

Qy    671 tcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcg 730
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3770 TCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 3829

Qy    731 caggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgg 790
          |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db   3830 CAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 3889

Qy    791 gtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaac 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3890 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 3949
```

```
Qy    851 tcaaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcac 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3950 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 4009

Qy    911 aaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggag 970
          |||||||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||
Db   4010 AAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 4069

Qy    971 ctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcaga 1030
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   4070 CTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGA 4129

Qy   1031 tgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcgg 1090
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4130 TGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG 4189

Qy   1091 catacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagc 1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4190 CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 4249

Qy   1151 gttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtga 1210
          |||||||||||||||||||||||||||||||||||||||||||||| || ||||| |
Db   4250 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCA 4309

Qy   1211 aactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccga 1270
          |||||||| || ||||||||||||||    || ||||||||||||||||||||||||||||
Db   4310 AACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 4369

Qy   1271 attttgaaaaagacgtaaaatacgatacg 1299
          ||||||| ||  |||| ||||| ||||||
Db   4370 ATTTTGAGAAGCACGTGAAATATGATACG 4398


RESULT 10
AX239715
LOCUS       AX239715               4425 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 17 from Patent WO0164920.
ACCESSION   AX239715
VERSION     AX239715.1  GI:15797369
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 4425)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 17 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .4425
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="deltaG983-ORF46.1"
BASE COUNT     1165 a   1259 c   1219 g    782 t
ORIGIN


  Query Match             70.4%;  Score 1176.2;  DB 6;  Length 4425;
  Best Local Similarity   97.3%;  Pred. No. 1.8e-296;
  Matches 1196;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy     71 cctcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaac 130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3170 CCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 3229

Qy    131 ccgacgggaaataccacctattcggcagcaggggggaacttgccgagcgcagcggtcata 190
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   3230 CCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATA 3289

Qy    191 tcggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggcca 250
          |||||||||||||| |||||||||||||||||||||||||| | || || ||||||||||
Db   3290 TCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCA 3349

Qy    251 ttaaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccc 310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3350 TTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCC 3409

Qy    311 ccttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggat 370
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3410 CCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGAT 3469

Qy    371 tcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacg 430
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3470 TTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACG 3529

Qy    431 ggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgaca 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3530 GGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACA 3589

Qy    491 taaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaac 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3590 TAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 3649

Qy    551 ggcttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggat 610
          |||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   3650 GGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGAT 3709

Qy    611 tcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctt 670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   3710 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCT 3769

Qy    671 tcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcg 730
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   3770 TCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 3829

Qy    731 caggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgg 790
          ||||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||
Db   3830 CAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 3889

Qy    791 gtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaac 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3890 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 3949

Qy    851 tcaaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcac 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3950 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 4009

Qy    911 aaggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggag 970
          |||||||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||
Db   4010 AAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 4069

Qy    971 ctgttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcaga 1030
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   4070 CTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGA 4129

Qy   1031 tgggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcgg 1090
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4130 TGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG 4189

Qy   1091 catacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagc 1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4190 CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 4249

Qy   1151 gttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtga 1210
          |||||||||||||||||||||||||||||||||||||||||||||| || ||||| |
Db   4250 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCA 4309

Qy   1211 aactggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccga 1270
          |||||||| || ||||||||||||||    || ||||||||||||||||||||||||||||
Db   4310 AACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 4369

Qy   1271 attttgaaaaagacgtaaaatacgatacg 1299
          ||||||| ||  |||| ||||| ||||||
Db   4370 ATTTTGAGAAGCACGTGAAATATGATACG 4398
```

```
RESULT 11
AX236447
LOCUS       AX236447               2019 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 140 from Patent WO0164922.
ACCESSION   AX236447
VERSION     AX236447.1  GI:15796045
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2019)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 140 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2019
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-741"
BASE COUNT      555 a    571 c    531 g    362 t
ORIGIN

  Query Match             70.4%;  Score 1175.6;  DB 6;  Length 2019;
  Best Local Similarity   97.2%;  Pred. No. 2.2e-296;
  Matches 1196;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgattctttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy     133 gacgggaaataccacctattcggcagcagggggggaacttgccgagcgcagcggtcatatc 192
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy     193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
           |||||||||| |||||||||||||||||||||||||| | || || ||||||||||||
Db     124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy     253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy     313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy     553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
           |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
           |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db     664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy     913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
           |||||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db     844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy     973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy    1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy    1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
           |||||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db    1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy    1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
           |||||| || |||||||||||||||    || ||||||||||||||||||||||||||||
Db    1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy    1273 tttgaaaaagacgtaaaatacgatacgaga 1302
           ||||| ||  |||| ||||| |||||| ||
Db    1204 TTTGAGAAGCACGTGAAATATGATACGGGA 1233


RESULT 12
AX239733
LOCUS       AX239733               2019 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 35 from Patent WO0164920.
ACCESSION   AX239733
VERSION     AX239733.1  GI:15797377
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2019)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 35 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2019
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-741"
BASE COUNT      555 a    571 c    531 g    362 t
ORIGIN

  Query Match             70.4%;  Score 1175.6;  DB 6;  Length 2019;
  Best Local Similarity   97.2%;  Pred. No. 2.2e-296;
  Matches 1196;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgattctttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy    133 gacgggaaataccacctattcggcagcaggggggaacttgccgagcgcagcggtcatatc 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy    193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
          |||||||||||| |||||||||||||||||||||||| | || || ||||||||||||
Db    124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy    253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy    313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy    553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
          |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
          |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db    664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy    913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
          |||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db    844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy    973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy   1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy   1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
          |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db   1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy   1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
          |||||| || |||||||||||||||||    || ||||||||||||||||||||||||||||
Db   1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy   1273 tttgaaaaagacgtaaaatacgatacgaga 1302
          ||||| ||  |||| ||||| |||||| ||
Db   1204 TTTGAGAAGCACGTGAAATATGATACGGGA 1233


RESULT 13
AX236451
LOCUS       AX236451               2256 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 144 from Patent WO0164922.
ACCESSION   AX236451
VERSION     AX236451.1  GI:15796047
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 144 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-961c"
BASE COUNT      698 a    608 c    545 g    405 t
ORIGIN


  Query Match             70.4%;  Score 1175.6;  DB 6;  Length 2256;
  Best Local Similarity   97.2%;  Pred. No. 2.2e-296;
  Matches 1196;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy    133 gacgggaaataccacctattcggcagcaggggggaacttgccgagcgcagcggtcatatc 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy    193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
          |||||||||||| |||||||||||||||||||||||| | || || ||||||||||||
Db    124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy    253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy    313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy    553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
          |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
```

```
Db    484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
          |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||||
Db    664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy    913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
          |||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db    844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy    973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy   1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy   1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
          |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db   1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy   1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
          |||||| || |||||||||||||||    || ||||||||||||||||||||||||||||
Db   1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy   1273 tttgaaaaagacgtaaaatacgatacgaga 1302
          ||||| ||  |||| ||||| |||||| ||
Db   1204 TTTGAGAAGCACGTGAAATATGATACGGGA 1233
```

```
RESULT  14
AX239737
LOCUS       AX239737               2256 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 39 from Patent WO0164920.
ACCESSION   AX239737
VERSION     AX239737.1  GI:15797379
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 39 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-961c"
BASE COUNT      698 a    608 c    545 g    405 t
```

ORIGIN

```
  Query Match             70.4%;  Score 1175.6;  DB 6;  Length 2256;
  Best Local Similarity   97.2%;  Pred. No. 2.2e-296;
  Matches 1196;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy     133 gacgggaaataccacctattcggcagcagggggggaacttgccgagcgcagcggtcatatc 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy     193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
           |||||||||| |||||||||||||||||||||||||| | || || ||||||||||||
Db     124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy     253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy     313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy     553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
           |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db     604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
           |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||||
Db     664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy     913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
           |||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db     844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy     973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy    1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy   1093 tacgccaaatacccgtcccctt-accattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
          |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db   1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy   1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
          |||||| || ||||||||||||||||    || ||||||||||||||||||||||||||
Db   1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy   1273 tttgaaaaagacgtaaaatacgatacgaga 1302
          ||||| ||  |||| ||||| |||||| ||
Db   1204 TTTGAGAAGCACGTGAAATATGATACGGGA 1233


RESULT 15
AX236449
LOCUS       AX236449               2421 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 142 from Patent WO0164922.
ACCESSION   AX236449
VERSION     AX236449.1  GI:15796046
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2421)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 142 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2421
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-961"
BASE COUNT      730 a    659 c    591 g    441 t
ORIGIN


  Query Match           70.4%;  Score 1175.6;  DB 6;  Length 2421;
  Best Local Similarity   97.2%;  Pred. No. 2.2e-296;
  Matches 1196;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy    133 gacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcatatc 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy    193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
          |||||||||||||||||||||||||||||||||||||| | || || ||||||||||||
Db    124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy    253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy    313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy    553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
          |||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
          |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||
Db    664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy    913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
          |||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db    844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy    973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy   1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy   1093 tacgccaaatacccgtcccctt-accattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
          |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db   1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy   1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
          |||||| || ||||||||||||||||    || ||||||||||||||||||||||||||
Db   1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy   1273 tttgaaaaagacgtaaaatacgatacgaga 1302
          ||||| ||  |||| ||||| |||||| ||
Db   1204 TTTGAGAAGCACGTGAAATATGATACGGGA 1233


RESULT 16
AX239735
LOCUS       AX239735               2421 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 37 from Patent WO0164920.
ACCESSION   AX239735
VERSION     AX239735.1  GI:15797378
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2421)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
```

```
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 37 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2421
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-961"
BASE COUNT      730 a    659 c    591 g    441 t
ORIGIN
```

Query Match 70.4%; Score 1175.6; DB 6; Length 2421;
Best Local Similarity 97.2%; Pred. No. 2.2e-296;
Matches 1196; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy      73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy     133 gacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcatatc 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy     193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
           |||||||||||| |||||||||||||||||||||||| | || || ||||||||||||
Db     124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy     253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy     313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy     553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
           |||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
           |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db     664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy     913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
           |||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db     844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy     973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy    1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy    1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
           |||||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db    1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy    1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
           |||||| || ||||||||||||||||    || ||||||||||||||||||||||||||||
Db    1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy    1273 tttgaaaaagacgtaaaatacgatacgaga 1302
           ||||| ||  |||| ||||| |||||| ||
Db    1204 TTTGAGAAGCACGTGAAATATGATACGGGA 1233
```

```
RESULT 17
AX236465
LOCUS       AX236465               2304 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 158 from Patent WO0164922.
ACCESSION   AX236465
VERSION     AX236465.1  GI:15796054
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2304)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 158 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2304
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961cL-ORF46.1"
BASE COUNT      710 a    619 c    555 g    420 t
ORIGIN
```

Query Match 70.3%; Score 1174.4; DB 6; Length 2304;
Best Local Similarity 97.1%; Pred. No. 4.6e-296;
Matches 1196; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy      73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1069 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1128

Qy     133 gacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcatatc 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1129 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1188

Qy     193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
           |||||||||||| |||||||||||||||||||||||| | || || ||||||||||||
Db    1189 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1248

Qy     253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1249 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1308
```

```
Qy    313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1309 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1368

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1369 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1428

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1429 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1488

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1489 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1548

Qy    553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
          |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1549 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1608

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1609 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1668

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1669 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1728

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
          |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||||
Db   1729 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1788

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1789 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1848

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1849 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 1908

Qy    913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
          |||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||||||
Db   1909 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 1968

Qy    973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1969 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 2028

Qy   1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2029 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2088

Qy   1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2089 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2148

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
          |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db   2149 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2208

Qy   1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
          |||||| || ||||||||||||||||    || |||||||||||||||||||||||||||||
Db   2209 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2268

Qy   1273 tttgaaaaagacgtaaaatacgatacgagaat 1304
          ||||| ||  |||| ||||| ||||||  | |
Db   2269 TTTGAGAAGCACGTGAAATATGATACGTAACT 2300
```

RESULT 18
AX239751
LOCUS AX239751 2304 bp DNA linear PAT 26-SEP-2001
DEFINITION Sequence 53 from Patent WO0164920.

ACCESSION AX239751
VERSION AX239751.1 GI:15797386
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 2304)
AUTHORS Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M. and Pizza,M.
TITLE Hybrid expression of neisserial proteins
JOURNAL Patent: WO 0164920-A 53 07-SEP-2001;
Chiron Spa (IT)
FEATURES Location/Qualifiers
source 1. .2304
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="961cL-ORF46.1"
BASE COUNT 710 a 619 c 555 g 420 t
ORIGIN

Query Match 70.3%; Score 1174.4; DB 6; Length 2304;
Best Local Similarity 97.1%; Pred. No. 4.6e-296;
Matches 1196; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy     73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1069 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1128

Qy    133 gacgggaaataccacctattcggcagcagggggggaacttgccgagcgcagcggtcatatc 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1129 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1188

Qy    193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
          |||||||||| |||||||||||||||||||||||||| | || || ||||||||||||
Db   1189 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1248

Qy    253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1249 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1308

Qy    313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1309 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1368

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1369 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1428

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1429 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1488

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1489 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1548

Qy    553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
          |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1549 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1608

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1609 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1668

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1669 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1728

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
          |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||||
Db   1729 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1788
```

```
Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1789 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1848

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1849 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 1908

Qy    913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
          |||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db   1909 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 1968

Qy    973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1969 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 2028

Qy   1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2029 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2088

Qy   1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2089 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2148

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
          |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db   2149 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2208

Qy   1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
          |||||| || |||||||||||||||    || ||||||||||||||||||||||||||||
Db   2209 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2268

Qy   1273 tttgaaaaagacgtaaaatacgatacgagaat 1304
          ||||| ||  |||| ||||| ||||||  | |
Db   2269 TTTGAGAAGCACGTGAAATATGATACGTAACT 2300


RESULT 19
AX236459
LOCUS       AX236459               2256 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 152 from Patent WO0164922.
ACCESSION   AX236459
VERSION     AX236459.1  GI:15796051
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 152 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961c-ORF46.1"
BASE COUNT      698 a    608 c    545 g    405 t
ORIGIN


  Query Match             70.3%;  Score 1174.2;  DB 6;  Length 2256;
  Best Local Similarity   97.3%;  Pred. No. 5.1e-296;
  Matches 1194;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy     73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1003 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1062

Qy    133 gacgggaaataccacctattcggcagcaggggggaacttgccgagcgcagcggtcatatc 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1063 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1122

Qy    193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
          |||||||||| |||||||||||||||||||||||||| | || || ||||||||||||
Db   1123 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1182

Qy    253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1183 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1242

Qy    313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1243 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1302

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1303 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1362

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1363 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1422

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1423 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1482

Qy    553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
          |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1483 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1542

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1543 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1602

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1603 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1662

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
          |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||
Db   1663 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1722

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1782

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1783 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 1842

Qy    913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
          |||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db   1843 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 1902

Qy    973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1903 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 1962

Qy   1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1963 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2022

Qy   1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2023 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2082

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
          |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db   2083 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2142

Qy   1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
          |||||| || |||||||||||||||    || ||||||||||||||||||||||||||||
Db   2143 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2202
```

```
Qy   1273 tttgaaaaagacgtaaaatacgatacg 1299
          ||||| ||  |||| ||||| ||||||
Db   2203 TTTGAGAAGCACGTGAAATATGATACG 2229


RESULT  20
AX239745
LOCUS       AX239745               2256 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 47 from Patent WO0164920.
ACCESSION   AX239745
VERSION     AX239745.1  GI:15797383
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 47 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961c-ORF46.1"
BASE COUNT      698 a    608 c    545 g    405 t
ORIGIN


  Query Match             70.3%;  Score 1174.2;  DB 6;  Length 2256;
  Best Local Similarity   97.3%;  Pred. No. 5.1e-296;
  Matches 1194;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy     73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1003 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1062

Qy    133 gacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcatatc 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1063 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1122

Qy    193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
          |||||||||||| |||||||||||||||||||||||| | || || ||||||||||||
Db   1123 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1182

Qy    253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1183 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1242

Qy    313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1243 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1302

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1303 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1362

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1363 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1422

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1423 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1482

Qy    553 cttgtcgaccgtttccacaatacggtagtatgctgacgcaaggagtaggcgacggattc 612
          |||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1483 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1542

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1543 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1602

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1603 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1662

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
          |||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||||
Db   1663 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1722

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1782

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1783 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 1842

Qy    913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
          |||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db   1843 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 1902

Qy    973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1903 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 1962

Qy    1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1963 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2022

Qy    1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2023 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2082

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
           |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db   2083 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2142

Qy    1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
           |||||| || |||||||||||||||    || ||||||||||||||||||||||||||||
Db   2143 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2202

Qy    1273 tttgaaaaagacgtaaaatacgatacg 1299
           ||||| ||  |||| ||||| ||||||
Db   2203 TTTGAGAAGCACGTGAAATATGATACG 2229


RESULT  21
AX236453
LOCUS       AX236453               2421 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 146 from Patent WO0164922.
ACCESSION   AX236453
VERSION     AX236453.1  GI:15796048
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2421)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 146 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2421
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961-ORF46.1"
BASE COUNT      730 a    659 c    591 g    441 t
ORIGIN


  Query Match             70.3%;  Score 1174.2;  DB 6;  Length 2421;
  Best Local Similarity   97.3%;  Pred. No. 5.2e-296;
```

```
  Matches 1194;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1168 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1227

Qy     133 gacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcatatc 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1228 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1287

Qy     193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
           |||||||||||| |||||||||||||||||||||||||| | || || ||||||||||||
Db    1288 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1347

Qy     253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1348 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1407

Qy     313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1408 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1467

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1468 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1527

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1528 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1587

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1588 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1647

Qy     553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
           |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1648 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1707

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    1708 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1767

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1768 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1827

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
           |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db    1828 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1887

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1888 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1947

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1948 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 2007

Qy     913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
           |||||||||||||||||||||||||| ||||| ||||||| ||||||||||||||||||
Db    2008 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 2067

Qy     973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    2068 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 2127

Qy    1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2128 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2187

Qy    1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2188 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2247

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
           |||||||||||||||||||||||||||||||||||||||||||||||| || ||||| |||
Db    2248 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2307

Qy    1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
           ||||||| || ||||||||||||||||    || ||||||||||||||||||||||||||||
Db    2308 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2367

Qy    1273 tttgaaaaagacgtaaaatacgatacg 1299
           ||||| ||  |||| ||||| ||||||
Db    2368 TTTGAGAAGCACGTGAAATATGATACG 2394


RESULT  22
AX239739
LOCUS       AX239739                2421 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 41 from Patent WO0164920.
ACCESSION   AX239739
VERSION     AX239739.1  GI:15797380
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2421)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 41 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2421
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961-ORF46.1"
BASE COUNT      730 a    659 c    591 g    441 t
ORIGIN


  Query Match             70.3%;  Score 1174.2;  DB 6;  Length 2421;
  Best Local Similarity   97.3%;  Pred. No. 5.2e-296;
  Matches 1194;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgattcttttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1168 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1227

Qy     133 gacgggaaataccacctattcggcagcagggggaacttgccgagcgcagcggtcatatc 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1228 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1287

Qy     193 ggattgggaaacatacaaagccatcagttgggcaacctgttcatccagcaggcggccatt 252
           |||||||||||| |||||||||||||||||||||||||| | || || ||||||||||||
Db    1288 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1347

Qy     253 aaaggaaatatcggctacattgtccgcttttccgatcacgggcacgaagtccattccccc 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1348 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1407

Qy     313 ttcgacaaccatgcctcacattccgattctgatgaagccggtagtcccgttgacggattc 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1408 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1467

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1468 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1527

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1528 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1587

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1588 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1647

Qy    553 cttgtcgaccgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattc 612
          |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1648 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1707

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagctttc 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1708 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1767

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1768 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1827

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgttatgcacggcttgggt 792
          |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db   1828 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1887

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1888 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1947

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1948 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 2007

Qy    913 ggcatagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagct 972
          |||||||||||||||||||||||||| ||||| ||||||| |||||||||||||||||||
Db   2008 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 2067

Qy    973 gttcggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   2068 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 2127

Qy   1033 ggcgagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2128 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2187

Qy   1093 tacgccaaatacccgtcccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2188 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2247

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaa 1212
          |||||||||||||||||||||||||||||||||||||||||| ||-|||| |||
Db   2248 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2307

Qy   1213 ctggcaaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaat 1272
          |||||| || ||||||||||||||    || ||||||||||||||||||||||||||||
Db   2308 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2367

Qy   1273 tttgaaaaagacgtaaaatacgatacg 1299
          ||||| ||  |||| ||||| ||||||
Db   2368 TTTGAGAAGCACGTGAAATATGATACG 2394


RESULT 23
A96426
LOCUS       A96426                   683 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 459 from Patent WO9924578.
ACCESSION   A96426
VERSION     A96426.1  GI:6780105
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 459 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      176 a    202 c    171 g    131 t      3 others
ORIGIN


  Query Match             39.1%;  Score 653.4;  DB 6;  Length 683;
  Best Local Similarity   97.1%;  Pred. No. 7.1e-160;
  Matches  663;  Conservative    1;  Mismatches   19;  Indels    0;  Gaps    0;

Qy     43 gcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgattcttttatccgg 102
          |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||||
Db      1 GCAGTGTGCCTNCCGATGCATGCACACGCCTCANATTTGGCAAACGATTCTTTTATCCGG 60

Qy    103 caggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcggcagcagg 162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGG 120

Qy    163 ggggaacttgccgagcgcagcggtcatatcggattgggaaacatacaaagccatcagttg 222
          |||||||||||||||||     |||||||||||||||| ||||||||||||||||||||
Db    121 GGGGAACTTGCCGAGCGCCAGTCTCATATCGGATTGGGAAAAATACAAAGCCATCAGTTG 180

Qy    223 ggcaacctgttcatccagcaggcggccattaaaggaaatatcggctacattgtccgcttt 282
          ||||||||| | || || ||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCAACCTGATGATTCAACAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTT 240

Qy    283 tccgatcacgggcacgaagtccattccccccttcgacaaccatgcctcacattccgattct 342
          |||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db    241 TCCGATCACGGGCACGAAGTCCATTCCCCSTTCGACAACCATGCCTCACATTCCGATTCT 300

Qy    343 gatgaagccggtagtcccgttgacggattcagcctttaccgcatccattgggacggatac 402
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    301 GATGAAGCCGGTAGTCCCGTTGACGGATTTAGCCTTTACCGCATCCATTGGGACGGATAC 360

Qy    403 gaacaccatcccgccgacggctatgacgggccacagggcggcggctatcccgctcccaaa 462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAA 420

Qy    463 ggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaatatccgcctcaac 522
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    421 GGCGCGAGGGATATATACAGTTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAAC 480

Qy    523 ctgaccgacaaccgcagcaccggacaacggcttgtcgaccgtttccacaataccggtagt 582
          |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db    481 CTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGTAGT 540

Qy    583 atgctgacgcaaggagtaggcgacggattcaaacgcgccacccgatacagccccgagctg 642
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTG 600

Qy    643 gacagatcgggcaatgccgccgaagctttcaacggcactgcagatatcgtcaaaaacatc 702
          |||||||||||||||||||||||||| |||||||||||||||||||| |||||||||
Db    601 GACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTTAAAAACATC 660

Qy    703 atcggcgcggcaggagaaattgt 725
          |||||||| ||||||||||||||
Db    661 ATCGGCGCTGCAGGAGAAATTGT 683


RESULT 24
AE002393/c
LOCUS       AE002393                 11700 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 35 of 206 of
            the complete genome.
ACCESSION   AE002393 AE002098
VERSION     AE002393.1  GI:7225586
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 11700)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
```

```
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 11700)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11700
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            201. .974
                     /gene="NMB0364"
     CDS             201. .974
                     /gene="NMB0364"
                     /note="similar to GB:L06299 PID:293962 percent identity:
                     99.61; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="FrpC operon protein"
                     /protein_id="AAF40807.1"
                     /db_xref="GI:7225587"
                     /translation="MRPYATTIYQLFILFIGSVFTMTSCEPVNEQTSFNNPEPMTGFE
                     HTVTFDFQGTKMVIPYGYLARYTQNNATKWLSDTPGQDAYSINLIEISVYYKKTDQGW
                     VLEPYNQQNKAHFIQFLRDGLDSVDDIVIRKDACSLSTTMGERLLTYGVKKMPSAYPE
                     YEAYEDKRHIPENPYFHEFYYIKKGENPAIITHRNNRINQTEEDSYSTSVGSCINGFT
                     VRYYPFIREKQQLTQQELVGYHQQVEQLVQSFVNNSSKK"
     gene            991. .1266
                     /gene="NMB0365"
                     /note="This region contains a match to at least one other
                     gene that is not full length,and is not the result of a
                     sequencing artifact; similar to GB:L06299 PID:293963
                     SP:P55127 percent identity: 98.72; identified by sequence
                     similarity; putative"
     gene            complement(1353. .2486)
                     /gene="NMB0366"
     CDS             complement(1353. .2486)
                     /gene="NMB0366"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40808.1"
                     /db_xref="GI:7225588"
                     /translation="MFKKFKPVLLSFFALVFAFWLGTGIAYEINPRWFLSDTATEVPK
                     NPNAFVAKLARLFRNADRAVVIVKESIRTEENLAGTVDDGPLQSEKDYLALAIRLSRL
                     KEKAKWFHVTEQEHGKEVWLDYHIGEGGLVAVSLSQRSPEAFVNAEYLYRNDRPFSVN
                     VYGGTVHGENYETTGEYRVVWQPDGSVFDAAGRGKIGEDVYEHCLGCYQMAQVYLAKY
                     RDVANDEQKVWDFRKESNRIASDSRNSVFYQNMRELMPRGMKANSLVVGYDADGLPQK
                     VYWSFDNGKKRQSFEYYLKNGNLFIAQSSTVALKADGVTADMQTYHAQQTWYLDGGRI
                     VREEKQGDRLPDFPLNLENLEKEVRRYAEAAARRSGGRRDLSH"
     gene            complement(2550. .2936)
                     /gene="NMB0367"
     CDS             complement(2550. .2936)
                     /gene="NMB0367"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40809.1"
                     /db_xref="GI:7225589"
                     /translation="MCIMCELKNFRRNITCFEGYDENSFIGKWYDDGVWDDEEYWKLE
                     NDLIEVRKKYPYPMDIPRYVVIGIGTIIDFLMVPNWKLFEIKASPWLPDSVGINERYE
                     RLKTMLRYIFTEKDIVNVQFDYYNKK"
     gene            complement(3005. .3463)
                     /gene="NMB0368"
     CDS             complement(3005. .3463)
                     /gene="NMB0368"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40810.1"
                     /db_xref="GI:7225590"
                     /translation="MVDAVVKTPEFLPFTSVGIFRFGADITQYKNILETFMYEPPDEF
                     RTEYYESPDSNLLISVKKKNKIISIFCYQELYFMGVNIIGLNFEDFKQLFHYPKSYGV
                     DKYYLSNESYPTYVYEFDEIGVQAWEAKGKIVTIIAGGKDNYSTEPYYDE"
     gene            complement(3444. .3833)
                     /gene="NMB0369"
     CDS             complement(3444. .3833)
                     /gene="NMB0369"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40811.1"
                     /db_xref="GI:7225591"
                     /translation="MKAGSRLLSESGRTQTVRKTVVKPKPLKAYNLTVADWHTYFVKG
                     NQAETEGVWVHNSCPPKRTGSSKNEKHGDGGRSQISAESKIAELTNKIIPGMSKNERL
                     KIKQKIRNIAKNANRKTKGEEHGRRGR"
     gene            complement(3965. .4342)
                     /gene="NMB0370"
     CDS             complement(3965. .4342)
                     /gene="NMB0370"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40812.1"
                     /db_xref="GI:7225592"
                     /translation="MCEFKDIIRNVPYFEGYDENSFIGKWYDDGVWDDEEYWKLENDL
                     IEVRKKYPYPMDIPRYVVIGIGTIIDFLMVPNWKLFEIKASPWLPDSVGIHERYERFT
                     TMLRYIFTEKDIVNVRFDYYNKK"
     gene            complement(4364. .4771)
                     /gene="NMB0371"
     CDS             complement(4364. .4771)
                     /gene="NMB0371"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40813.1"
                     /db_xref="GI:7225593"
                     /translation="MCEFKDFRRNIPCFKEYDENSFIGKWYDDGVWDDEEYWKLENDL
                     IEVRKKYPYPMDIPRDIVIGIGTIIDFLMVQNWKLFEIKASTWLPKSVKINERYERFR
                     VMLRYIFTDLDVEDWKFFYFPIQHSKGRLRRKV"
     gene            complement(4777. .5454)
                     /gene="NMB0372"
     CDS             complement(4777. .5454)
```

```
                     /gene="NMB0372"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40814.1"
                     /db_xref="GI:7225594"
                     /translation="MVKTADGYKAIARIRTGDRVFAKDEASGKTGYKPVTARYGNPYQ
                     ETVYIEISDGIGNNQTLISNKIHPFYSQGKWIQAGRLKKGDTLLSESGAKQTVQNITL
                     KQQPLKAYNLTVADWHTYFVKGNRAETEGVWVHNDCPPRKTPSTPIYGNDSEAYAAAK
                     ELGYRKIKERTRNDAAIFKKGKSYISRDVDSHNGGAWKEASSPEKLNRKETRNGTFDK
                     NLNRIGD"
     gene            complement(5669. .6028)
                     /gene="NMB0373"
     CDS             complement(5669. .6028)
                     /gene="NMB0373"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40815.1"
                     /db_xref="GI:7225595"
                     /translation="MASYVSIKGWIECSDDDIKIIQENINNFWNNCPFNIEEKESAKI
                     YKSGWVFPTNSFNWSSYIFFGACVKSYFIIYFEKCIKTIMELDIEISGFFELDYYEDN
                     YKVNWKINNGNLIQTTN"
     gene            complement(6031. .7434)
                     /gene="NMB0374"
     CDS             complement(6031. .7434)
                     /gene="NMB0374"
                     /note="MafB-related protein; identified by Oxford;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="MafB-related protein"
                     /protein_id="AAF62308.1"
                     /db_xref="GI:7413426"
                     /translation="MKPLRRLTNLLAACAVAAAALIQPALAADLAQDPFITDNAQRQH
                     YEPGGKYHLFGDPRGSVSDRTGKINVIQDYTHQMGNLLIQQANINGTIGYHTRFSGHG
                     HEEHAPFDNHAADSASEEKGNVDEGFTVYRLNWEGHEHHPADAYDGPKGGNYPKPTGA
                     RDEYTYHVNGTARSIKLNPTDTRSIRQRISDNYSNLGSNFSDRADEANRKMFEHNAKL

  Query Match             17.6%;  Score 294.4;  DB 1;  Length 11700;
  Best Local Similarity   58.1%;  Pred. No. 1.1e-65;
  Matches  539;  Conservative    0;  Mismatches  386;  Indels    3;  Gaps    1;

Qy      35 ccatactggcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgattctt 94
           || ||  ||| |     ||      ||  || | |||  | || ||||  | ||  | |
Db    7391 CCGTAGCGGCGGCCGCACTCATACAGCCCGCCCTCGCGGCGGACTTGGCGCAAGACCCGT 7332

Qy      95 ttatccggcaggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcg 154
           | ||     |      | | || ||||| | |||||||| ||| ||||||||||| ||||
Db    7331 TCATTACCGATAACGCCCAACGGCAGCACTACGAACCCGGCGGCAAATACCACCTCTTCG 7272

Qy     155 gcagcaggggggaac---ttgccgagcgcagcggtcatatcggattgggaaacatacaaa 211
           ||  |  | | |      || |||| |||| |||  | |||    |     |     | |
Db    7271 GCGACCCGCGCGGCAGCGTTTCCGACCGCACCGGCAAAATCAACGTCATCCAAGACTATA 7212

Qy     212 gccatcagttgggcaacctgttcatccagcaggcggccattaaaggaaatatcggctaca 271
            ||| ||| |||||||||||| ||||||| |||||   ||| || || |  |||||||||
Db    7211 CCCACCAGATGGGCAACCTGCTCATCCAACAGGCAAACATCAACGGCACAATCGGCTACC 7152

Qy     272 ttgtccgcttttccgatcacgggcacgaagtccattccccccttcgacaaccatgcctcac 331
           ||||||||||||  ||||| |||||||  ||  |||||||||||||||| ||| |
Db    7151 ACACCCGCTTTTCCGGACACGGACACGAAGAACACGCCCCCTTCGACAACCACGCCGCCG 7092

Qy     332 attccgattctgatgaagccggtagtcccgttgacggattcagcctttaccgcatccatt 391
           |   ||     || ||| ||   || |      |  || || || | | | |||||  || | |
Db    7091 ACAGCGCGAGCGAAGAAAAAGGCAACGTTGACGAAGGCTTTACCGTATACCGGCTCAACT 7032

Qy     392 gggacggatacgaacaccatcccgccgacggctatgacgggccacagggcggcggctatc 451
           |||| ||| ||||||| |||||||||||  | ||| ||||| ||  ||||||||   || |

Db    7031 GGGAAGGACACGAACATCATCCCGCCGATGCCTACGACGGCCCGAAGGGCGGCAATTACC 6972

Qy     452 ccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaata 511
           ||   || |  ||||  | ||   |||| ||| || | || |||   ||||  | ||
Db    6971 CCAAACCTACGGGCGCACGAGACGAATACACCTATCACGTCAACGGCACAGCCCGCAGTA 6912

Qy     512 tccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgtcgaccgtttccaca 571
           ||   ||||| | |||||||| |||||||| | | |||||  |   ||||  || |  ||
Db    6911 TCAAACTCAATCCGACCGACACCCGCAGCATCCGGCAACGCATATCCGACAATTACAGCA 6852

Qy     572 ataccggtagtatgctgacgcaaggagtaggcgacggattcaaacgcgccacccgataca 631
           |   ||| || |   |  |  |  | |  |  || |    || |       |   |||
Db    6851 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 6792

Qy     632 gccccgagctggacagatcgggcaatgccgccgaagctttcaacggcactgcagatatcg 691
           || |||| ||| | | ||||||   |   ||   | ||||||||   || |    ||
Db    6791 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 6732

Qy     692 tcaaaaacatcatcggcgcggcaggagaaattgtcggcgcaggcgatgccgtgcagggta 751
           |||  | |   |   || || |||    | |||  ||||||     || | || |
Db    6731 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCTTGGGCATAGGCGACATACTGTACGGAA 6672

Qy     752 taagcgaaggctcaaacattgctgttatgcacggcttgggtctgctttccaccgaaaaca 811
           || | | |  | |||  || |  |||| |  | | |  |   |  || ||||   ||
Db    6671 CGCGCTATGCCATAGACAAAGCCGCAATGCGCAACATCGCCCCCTTGCCCGCCGAGGGCA 6612

Qy     812 agatggcgcgcatcaacgatttggcagatatggcgcaactcaaagactatgccgcagcag 871
           |  | ||   ||||  ||  ||||      |||||    |  || |  || |    | ||
Db    6611 AATTCGCCGTCATCGGCGGCTTGGGCAGCGTGGCGGGCTTTGAAAAGAATACGCGCGAAG 6552

Qy     872 ccatccgcgattgggcagtccaaaaccccaatgccgcacaaggcatagaagccgtcagca 931
           || |  |   |||  |    |||| ||||||||||  ||  | | |||||||||  ||
Db    6551 CCGTTGACCGGTGGATACAGGAAAATCCCAATGCCGCCGAAACCGTCGAAGCCGTCTTCA 6492

Qy     932 atatctttacggcagtcatccccgtcaa 959
           |  |     | |||| ||    ||  ||
Db    6491 ACGTTGCCGCAGCAGCCAAAGTCGCGAA 6464


RESULT  25
AX044029/c
LOCUS       AX044029               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 108 from Patent WO0066791.
ACCESSION   AX044029
VERSION     AX044029.1  GI:11342913
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 108 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    84410 a  84863 c  94187 g  86520 t
ORIGIN


  Query Match             17.6%;  Score 294.4;  DB 6;  Length 349980;
```

```
  Best Local Similarity   58.1%;  Pred. No. 2.3e-65;
  Matches  539;  Conservative    0;  Mismatches  386;  Indels    3;  Gaps    1;

Qy      35 ccatactggcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgattctt 94
           || ||  ||| |    ||       ||  || | |||  | || |||||  | ||  | |
Db   77278 CCGTAGCGGCGGCCGCACTCATACAGCCCGCCCTCGCGGCGGACTTGGCGCAAGACCCGT 77219

Qy      95 ttatccggcaggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcg 154
           | ||     |       | | || ||||| | |||||||| ||| ||||||||||| ||||
Db   77218 TCATTACCGATAACGCCCAACGGCAGCACTACGAACCCGGCGGCAAATACCACCTCTTCG 77159

Qy     155 gcagcaggggggaac---ttgccgagcgcagcggtcatatcggattgggaaacatacaaa 211
           ||  |  | | |      || |||| |||| |||  | |||     |      |     | |
Db   77158 GCGACCCGCGCGGCAGCGTTTCCGACCGCACCGGCAAAATCAACGTCATCCAAGACTATA 77099

Qy     212 gccatcagttgggcaacctgttcatccagcaggcggccattaaaggaaatatcggctaca 271
           ||| ||| |||||||||| |||||| |||||   ||| || || |  |||||||||
Db   77098 CCCACCAGATGGGCAACCTGCTCATCCAACAGGCAAACATCAACGGCACAATCGGCTACC 77039

Qy     272 ttgtccgcttttccgatcacgggcacgaagtccattccccccttcgacaaccatgcctcac 331
           ||||||||||  ||||| |||||||  ||  ||||||||||||||| ||| |
Db   77038 ACACCCGCTTTTCCGGACACGGACACGAAGAACACGCCCCCTTCGACAACCACGCCGCCG 76979

Qy     332 attccgattctgatgaagccggtagtcccgttgacggattcagcctttaccgcatccatt 391
           |   ||     || |||   || |     |  || || || | | | |||||  || | |
Db   76978 ACAGCGCGAGCGAAGAAAAAGGCAACGTTGACGAAGGCTTTACCGTATACCGGCTCAACT 76919

Qy     392 gggacggatacgaacaccatcccgccgacggctatgacgggccacagggcggcggctatc 451
           |||| ||| |||||||| |||||||||| | ||| ||||| ||  ||||||||   || |
Db   76918 GGGAAGGACACGAACATCATCCCGCCGATGCCTACGACGGCCCGAAGGGCGGCAATTACC 76859

Qy     452 ccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaata 511
           ||   || |  ||||| | ||   ||||| ||| || | || |||   ||||  | ||
Db   76858 CCAAACCTACGGGCGCACGAGACGAATACACCTATCACGTCAACGGCACAGCCCGCAGTA 76799

Qy     512 tccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgtcgaccgtttccaca 571
           ||   ||||| | |||||||| |||||||| | | |||||  |   ||||  || |  ||
Db   76798 TCAAACTCAATCCGACCGACACCCGCAGCATCCGGCAACGCATATCCGACAATTACAGCA 76739

Qy     572 ataccggtagtatgctgacgcaaggagtaggcgacggattcaaacgcgccacccgataca 631
           |   ||| || |   |  |  | |  |  |  || |     |    |||
Db   76738 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 76679

Qy     632 gccccgagctggacagatcgggcaatgccgccgaagctttcaacggcactgcagatatcg 691
           || |||| ||| | | ||||||   |   ||   | ||||||||   || |    ||
Db   76678 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 76619

Qy     692 tcaaaaacatcatcggcgcggcaggagaaattgtcggcgcaggcgatgccgtgcagggta 751
           |||  | |   |  || || |||    | |||  ||||||     || | || |
Db   76618 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCTTGGGCATAGGCGACATACTGTACGGAA 76559

Qy     752 taagcgaaggctcaaacattgctgttatgcacggcttgggtctgctttccaccgaaaaca 811
           || | | |  | |||  || |  |||| |  | | |  |   || ||||   ||
Db   76558 CGCGCTATGCCATAGACAAAGCCGCAATGCGCAACATCGCCCCCTTGCCCGCCGAGGGCA 76499

Qy     812 agatggcgcgcatcaacgatttggcagatatggcgcaactcaaagactatgccgcagcag 871
           |  | ||   ||||  ||  ||||      |||||    |  || |  || |    | ||
Db   76498 AATTCGCCGTCATCGGCGGCTTGGGCAGCGTGGCGGGCTTTGAAAAGAATACGCGCGAAG 76439

Qy     872 ccatccgcgattgggcagtccaaaaccccaatgccgcacaaggcatagaagccgtcagca 931
           || |   |  |||  |    |||| ||||||||||  ||  | | |||||||||  ||
Db   76438 CCGTTGACCGGTGGATACAGGAAAATCCCAATGCCGCCGAAACCGTCGAAGCCGTCTTCA 76379

Qy     932 atatctttacggcagtcatccccgtcaa 959
           |  |     | |||| ||    ||  ||
Db   76378 ACGTTGCCGCAGCAGCCAAAGTCGCGAA 76351


RESULT  26
NMA7Z2491
LOCUS       NMA7Z2491              195767 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 7/7.
ACCESSION   AL162758 AL157959
VERSION     AL162758.2  GI:7380672
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 195767)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 195767)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .195767
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            80. .952
                     /gene="NMA2045"
     CDS             80. .952
                     /gene="NMA2045"
                     /note="NMA2045, possible oxidoreductase, len: 290aa;
                     similar to many eg. SW:P20692 (TYRA_BACSU) prephenate
                     dehydrogenase from Bacillus subtilis (372 aa) fasta
                     scores; E(): 2.8e-30, 39.4% identity in 284 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="putative oxidoreductase"
                     /protein_id="CAB85263.1"
                     /db_xref="GI:7380673"
                     /db_xref="SPTREMBL:Q9JT12"
                     /translation="MPILNHIALIGVGLIGGSFVLDLKRQGLVRTVTGIDTDRDNLER
                     ALERGVIDRASVAIDADSIGGADLVLIATPVATVPAILTALRPVLPEHTWISDVGSTK
                     SSVIEAFRRCLPDRLHRCIAAHPIAGSDRSGAQAAQFGLFRHRKLIITPHGGEHSDGI
                     ALIENLWHAVGAEIYTMDAQRHDAVFAAVSHMPHLTAFAYVHQILDHPDGQEYLKFAA
                     TGFRDFTRIASGHPAVWADICLANKDSLLQLVQGLGKQLDVLADILTADDREALYRYF
                     EEAKTTRDRWLDGN"
     misc_feature    complement(547. .556)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    960. .969
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(1009. .2211)
                     /gene="NMA2046"
     CDS             complement(1009. .2211)
                     /gene="NMA2046"
                     /note="NMA2046, hypothetical integral membrane protein,
                     len: 400aa; similar to TR:P95556 (EMBL:Z73914)
                     hypothetical protein from Pseudomonas stutzeri (396 aa)
                     fasta scores; E(): 0, 35.0% identity in 391 aa overlap.
                     Contains possible membrane-spanning hydrophobic regions."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical integral membrane protein"
                     /protein_id="CAB85264.1"
```

```
                     /db_xref="GI:7380674"
                     /db_xref="SPTREMBL:Q9JT11"
                     /translation="MIPFCGKYINDKQGTHMKFTKHPVWAMAFRPFYSLAALYGALSV
                     LLWGFGYTGTHELSGFYWHAHEMIWGYAGLVVIAFLLTAVATWTGQPPTRGGVLVGLT
                     IFWLAARIAAFIPGWGASASGILGTLFFWYGAVCMALPVIRSQNQRNYVAVFALFVLG
                     GTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLIGTRIISFFTSKRLNVPQIPSPK
                     WVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFTVQVYRWWYKPVLKEPMLWIL
                     FAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLTLGMMARTALGHTGNPIYP
                     PPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFALALLVYAWKYIPWLIR
                     PRSDGRPG"
gene                 complement(2208. .2576)
                     /gene="NMA2047"
CDS                  complement(2208. .2576)
                     /gene="NMA2047"
                     /note="NMA2047, hypothetical protein, len: 122aa; weakly
                     similar to SW:P71932 (YQ03_MYCTU) hypothetical protein
                     from Mycobacterium tuberculosis (161 aa) fasta scores;
                     E(): 0.017, 32.5% identity in 117 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB85265.1"
                     /db_xref="GI:7380675"
                     /db_xref="SPTREMBL:Q9JT10"
                     /translation="MKPLKRHPALIGLSRDHHHSLSLCVRLLRTPEERHRDELEPHFS
                     ELETHFREEEAKFAPIWQNVAPELKQRFEKDHARLRQMMASPEYGNAAWNTAFATTLR
                     DHARFEERELFPAVEPFLPA"
misc_feature         complement(2622. .2631)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
gene                 2750. .3193
                     /gene="NMA2048"
CDS                  2750. .3193
                     /gene="NMA2048"
                     /note="NMA2048, conserved hypothetical protein, len:
                     147aa; similar to many eg. SW:P21498 (YJEB_ECOLI)
                     hypothetical protein from Escherichia coli (141 aa) fasta
                     scores; E(): 5.9e-23, 47.9% identity in 140 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAB85266.1"
                     /db_xref="GI:7380676"
                     /db_xref="SPTREMBL:Q9JT09"
                     /translation="MYLTQHTDYGLRVLIYTAINDDALVNISTIAVTYGISKSHLMKV
                     VTALVKGGFLHSVRGKGGGLRLASPPERINIGAVVRHLEPMQLVECMGENNECLITPS
                     CRLTGILGGAMKSFFTYLDGFTLQDLLNKPTYDLLYEPRIPIAVQ"
stem_loop            3275. .3314
                     /note="Stem loop containing DNA uptake sequences: gcaat
                     gccgtctgaa cagccgatgc ttcagacggc atcgg"
misc_feature         3280. .3289
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
misc_feature         complement(3300. .3309)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
gene                 complement(3320. .3940)
                     /gene="NMA2049"
CDS                  complement(3320. .3940)
                     /gene="NMA2049"
                     /note="NMA2049, conserved hypothetical protein, len:
                     206aa; similar to many eg. SW:P45847 (YCIO_ECOLI)
                     hypothetical protein from Escherichia coli (206 aa) fasta
                     scores; E(): 0, 62.9% identity in 205 aa overlap. Contains
                     Pfam match to entry PF01300 Sua5_yciO_yrdC, SUA5/yciO/yrdC
                     family and Prosite match to PS01147 SUA5/yciO/yrdC family
                     signature."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAB85267.1"
                     /db_xref="GI:7380677"
                     /db_xref="SPTREMBL:Q9JT08"
                     /translation="MAQFFAIHPDNPQERLIKQAVEIVNKGGVVVYPTDSCYALGCKL
                     GDKAAMERILSIRKIDLKHHLTLMCADLSELGTYAKIDNVQFRQLKAATPGPYTFILQ
                     ATKDVPARTLHPKRKTIGLRIPDNAIAQALLGELGEPLLSCTLMLPEDGEPLTDPYEI
                     RECLEHAVDLVIDGGWCGTEPTTVVDMTDGTELVRQGCGDTAVFGL"
misc_feature         complement(3350. .3877)
                     /gene="NMA2049"
                     /note="Pfam match to entry PF01300 Sua5_yciO_yrdC,
                     SUA5/yciO/yrdC family, score 192.00, E-value 9.6e-54"
misc_feature         complement(3818. .3856)
                     /gene="NMA2049"
                     /note="PS01147 SUA5/yciO/yrdC family signature."
repeat_unit          4027. .4046
                     /note=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
                     /label=dRS3
repeat_unit          4047. .4102
                     /note="85% identity to consensus
                     CCAGTCCGTTCAGTTTCAGTCATTTCCGATAAATTCCTGTTGCTTTTCATTTCTAG"
                     /label=RS17
misc_feature         complement(4262. .4271)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
gene                 complement(4298. .5497)
                     /gene="ackA2"
CDS                  complement(4298. .5497)
                     /gene="ackA2"
                     /EC_number="2.7.2.1"
                     /note="NMA2050, ackA2, acetate kinase, len: 399aa; similar
                     to many eg. SW:P15046 (ACKA_ECOLI) ackA, acetate kinase
                     from Escherichia coli (400 aa) fasta scores; E(): 0, 49.5%
                     identity in 400 aa overlap. Also similar to NMA1718 from
                     Neisseria meningitidis fasta scores; E(): 0, 63.3%
                     identity in 398 aa overlap. Contains Pfam match to entry
                     PF00871 Acetate_kinase, Acetokinase family; Prosite match
                     to PS01075 Acetate and butyrate kinases family signature 1
                     and Prosite match to PS01076 Acetate and butyrate kinases
                     family signature 2."
                     /codon_start=1
                     /transl_table=11
                     /product="acetate kinase"
                     /protein_id="CAB85268.1"
                     /db_xref="GI:7380678"
                     /db_xref="SPTREMBL:Q9JT07"
                     /translation="MSDQLILVLNCGSSSLKGAVIDRKSGSVVLSCLGERLTTPEAVI
                     TFSKDGNKRQVPLSGRNCHAGAVGMLLNELEKHELHDRIQAVGHRIAHGGEKYSESVL
                     IDQAVMDELNACIPLAPLHNPANISGILAAQEHFPGLPNVGVMDTSFHQTMPERAYTY
                     AVPRELRKKYAFRRYGFHGTSMRYVAPEAACILGKPLEDIRMIIAHLGNGASITAIKN
                     GKSVDTSMGFTPIEGLVMGTRCGDIDPGVYSYLTSHAGLDVAQVDEMLNKKSGLLGIS
                     ELSNDCRTLEIAADEGHEGARLALEVMTYRLAKYIASMAVGCGGVDALVFTGGIGENS
                     RNIRAKTVSYLDFLGLHIDTKANMEKRYGNSGIISPTDSSPAVLVVPTNEELMIACDT
                     AELVGIL"
misc_feature         complement(4328. .5485)
```

```
  Query Match             17.1%;  Score 286;  DB 1;  Length 195767;
  Best Local Similarity   57.7%;  Pred. No. 3.1e-63;
  Matches  530;  Conservative    0;  Mismatches  385;  Indels    3;  Gaps    1;

Qy      35 ccatactggcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgattctt 94
           || ||  ||| |     ||      ||  || | |||  | || |||||  | ||  | |
Db   79040 CCGTAGCGGCGGCCGCACTCATACAGCCCGCCCTCGCGGCGGACTTGGCGCAAGACCCGT 79099

Qy      95 ttatccggcaggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcg 154
           | ||     |       | | || ||||| | ||||||||  || ||||||||||| ||||
Db   79100 TCATTACCGATAACGCCCAACGGCAGCACTACGAACCCGGAGGCAAATACCACCTCTTCG 79159

Qy     155 gcagcagggggggaacttg---ccgagcgcagcggtcatatcggattgggaaacatacaaa 211
           ||  |  | | |     |   |||| |||| |||||| |||    |      |     | |
Db   79160 GCGACCCGCGCGGCAGCGTCTCCGACCGCACCGGTCAAATCAACGTCATCCAAGACTATA 79219

Qy     212 gccatcagttgggcaacctgttcatccagcaggcggccattaaaggaaatatcggctaca 271
             ||| | | |||||||||||| ||||||||||||||   ||| || || |  |||||||||
Db   79220 CCCACCGGATGGGCAACCTGCTCATCCAGCAGGCAAACATCAACGGCACAATCGGCTACC 79279

Qy     272 ttgtccgcttttccgatcacgggcacgaagtccattccccccttcgacaaccatgcctcac 331
```

```
Db  79280 ACACCCGCTTTTCCGGACACGGATACGAAGAACACGCCCCCTTCGACAACCACGCCGCCG 79339

Qy    332 attccgattctgatgaagccggtagtcccgttgacggattcagcctttaccgcatccatt 391
Db  79340 ACAGCGCGAGCGAAGAAAAAGGCAACGTTGACGAAGGCTTTACCGTATACCGGCTCAACT 79399

Qy    392 gggacggatacgaacaccatcccgccgacggctatgacgggccacagggcggcggctatc 451
Db  79400 GGGAAGGACACGAACATCATCCCGCCGATGCCTACGACGGCCCGAAGGGCGGCAATTACC 79459

Qy    452 ccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaata 511
Db  79460 CCAAACCTACGGGTGCACGCGACGAATACACCTATCACGTCAACGGCACAGCACGCAGCA 79519

Qy    512 tccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgtcgaccgtttccaca 571
Db  79520 TCAAACTCAATCCGACCGACACCCGCAGCATCCGGCAACGCATATCCGACAATTACAGCA 79579

Qy    572 ataccggtagtatgctgacgcaaggagtaggcgacggattcaaacgcgccacccgataca 631
Db  79580 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 79639

Qy    632 gccccgagctggacagatcgggcaatgccgccgaagctttcaacggcactgcagatatcg 691
Db  79640 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 79699

Qy    692 tcaaaaacatcatcggcgcggcaggagaaattgtcggcgcaggcgatgccgtgcagggta 751
Db  79700 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCTTGGGCATAGGCGACATACTGTACGGAA 79759

Qy    752 taagcgaaggctcaaacattgctgttatgcacggcttgggtctgctttccaccgaaaaca 811
Db  79760 CGCGCTATGCCATAGACAAAGCCGCAATGCGCAACATCGCCCCCTTGCCCGCCGAGGGCA 79819

Qy    812 agatggcgcgcatcaacgatttggcagatatggcgcaactcaaagactatgccgcagcag 871
Db  79820 AATTCGCCGTCATCGGCGGCTTGGGCAGCGTGGCGGGCTTTGAAAAAAATACGCGCGAAG 79879

Qy    872 ccatccgcgattgggcagtccaaaaccccaatgccgcacaaggcatagaagccgtcagca 931
Db  79880 CCGTTGACCGGTGGATACAGGAAAACCCCAATGCCGCCGAAACCGTCGAAGCCCTGGTCA 79939

Qy    932 atatctttacggcagtca 949
Db  79940 ACGTCCTGCCGTTTGCCA 79957


RESULT  27
AE002420
LOCUS       AE002420               10057 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 62 of 206 of
            the complete genome.
ACCESSION   AE002420 AE002098
VERSION     AE002420.1  GI:7225876
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10057)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10057)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10057
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            126. .350
                     /gene="NMB0649"
     CDS             126. .350
                     /gene="NMB0649"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41070.1"
                     /db_xref="GI:7225877"
                     /translation="MRLATKDLAEAIRKGQVRKSSFNTEQLRAIEKGESKIPDYTWHH
                     HQDTGRMQLIREGLHHDTGHIGWEAMNKGR"
     gene            323. .742
                     /gene="NMB0650"
     CDS             323. .742
                     /gene="NMB0650"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41071.1"
                     /db_xref="GI:7225878"
                     /translation="MGSNEQRKVTMWKIIKEDSDDLEFAIKCLFSQSIDLNEFKLWIE
                     QVIRDMPIEDIPFYIFDLADFDGGIADIDNIVGFVSSCRLSKSKKNALTGIAFLRGID
                     VYDPPISKEKALKALEKHPEIYQKFQHFFPFVELPPL"
     gene            843. .1232
                     /gene="NMB0651"
     CDS             843. .1232
                     /gene="NMB0651"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41072.1"
                     /db_xref="GI:7225879"
                     /translation="MWKIIKEDSDDLGFAIKCLFSQSIDLNEFKLWIEQVIRDMPIED
                     IPFYIFDLADFDGGIADIDNIVGFVSSCRLSKSKKNALTGIAFLRGIDVYDPPISKEK
                     ALKALEKHPEIYQKFQHFFPFVELPPL"
     gene            1355. .2296
                     /gene="NMB0652"
     CDS             1355. .2296
                     /gene="NMB0652"
                     /note="similar to GP:4838554 percent identity: 79.35;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="mafA protein"
                     /protein_id="AAF41073.1"
                     /db_xref="GI:7225880"
```

```
                /translation="MKTLLLLIPLVLTACGTLTGIPAHGGGKRFAVEQELVAASSRAA
                VKEMDLSALKGRKAALYVSVMGDQGSGNISGGRYSIDALIRGGYHNNPESATQYSYPA
                YDTTATTKSDALSSVTTSTSLLNAPAAALTKNSGRKGERSAGLSVNGTGDYRNETLLA
                NPRDVSFLTNLIQTVFYLRGIEVVPPEYADTDVFVTVDVFGTVRSRTELHLYNAETLK
                AQTKLEYFAVDRDSRKLLITPKTAAYESQYQEQYALWTGPYKVSKTVKASDRLMVDFS
                DITPYGDTTAQNRPDFKQNNGKKPDVGNEVIRRRKGG"
gene            2300. .3568
                /gene="NMB0653"
CDS             2300. .3568
                /gene="NMB0653"
                /note="MafB-related protein; identified by Oxford;
                putative"
                /codon_start=1
                /transl_table=11
                /product="MafB-related protein"
                /protein_id="AAF62314.1"
                /db_xref="GI:7413436"
                /translation="MKPLRRLTNLLAACAVAAAALIQPALAADLAQDPFITDNAQRQH
                YEPGGKYHLFGDPRGSVSDRTGKINVIQDYTHQMGNLLIQQANINGTIGYHTRFSGHG
                HEEHAPFDNHAADSASEEKGNVDEGFTVYRLNWEGHEHHPADAYDGPKGGNYPKPTGA
                RDEYTYHVNGTARSIKLNPTDTRSIRQRISDNYSNLGSNFSDRADEANRKMFEHNAKL
                DRWGNSMEFINGVAAGALNPFISAGEAVDQWMQENPNAAETVEALVNVLPFAKVKNLT
                KAAKPGKAAVSGDFSDSYKHNTASRLSQSVDGEMFQTRNVDFKAKSIGTKIHDGAQGK
                HISGHRNYIEGKSTLNQNINPQELLNGIHSGAYPVISKGARGNPVVDFGYPIGSDGKS
                GLSTNFGTIHSGKNGVHIVPANPKTIKKVQ"
gene            3571. .3888
                /gene="NMB0654"
CDS             3571. .3888
                /gene="NMB0654"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF41074.1"
                /db_xref="GI:7225881"
                /translation="MNILPSWLRVGMNIAMLGMIHSDIRLITVDYEEGRRFLKIKNYL
                SREAITEDHEDMEYLITELWSMCGEYFDEADFECIYSNHSSMELNQINGAVFRRKELI
                SQA"
gene            4023. .4730
                /gene="NMB0655"
CDS             4023. .4730
                /gene="NMB0655"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF41075.1"
                /db_xref="GI:7225882"
                /translation="MVKTADGYKAIARIRTGDRVFAKDEASGKTGYKPVTARYGNPYQ
                ETVYIEISDGIGNNQTLISNKIHPFYSQGKWIQAGRLKKGDTLLSESGAKQTVQNITF
                KQQPLKAYNLTVADWHTYFVKGSQAETEGVWVHNDCPYDKGNQRYKDASYHGKNDNSV
                KSRAPTNGQAALDNSVQVKSTSPRRVGVDKANNEIVVLNKTQTFNNGSAEYHGHVRSW
                QDLHTDQKNALKKAGLD"
gene            4756. .5085
                /gene="NMB0656"
CDS             4756. .5085
                /gene="NMB0656"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF41076.1"
                /db_xref="GI:7225883"
                /translation="MTDKSKTEKLISSDDKQSVIDGILDMVFNSKAYEVPWISEKLME
                LSKNKDLDIAGLSLTCFGHLARLHSNIGDYDKVIPLLHSKQDDPELQGRAEDALEDIS
                LFLSENH"
gene            5180. .5383
                /gene="NMB0657"
CDS             5180. .5383
                /gene="NMB0657"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF41077.1"
                /db_xref="GI:7225884"
                /translation="MRTEFFKNPTLAPSTAAPWSKRQTAEKQTPSVVLPFADEYGSTP
                MPPKPSKPPSTLPPPKPQSWQKR"
gene            5211. .5471
                /gene="NMB0658"
CDS             5211. .5471
                /gene="NMB0658"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF62315.1"
                /db_xref="GI:7413437"
                /translation="MLLPRQHLGQNGRRLKSKHRPSCCRLRMSTGQPQCRRNRRSRLQ
                HCRRQSRKVGKNGKTGEIKADGRKVNVRIDSTEADLLYPAGQ"
gene            5502. .5903
                /gene="NMB0659"
CDS             5502. .5903
                /gene="NMB0659"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
```

Query Match 14.2%; Score 237.2; DB 1; Length 10057;
Best Local Similarity 59.7%; Pred. No. 9.7e-51;
Matches 417; Conservative 0; Mismatches 278; Indels 3; Gaps 1;

```
Qy      35 ccatactggcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgattctt 94
           || ||  ||| |     ||      ||  || | |||  | || |||||  | ||  | |
Db    2343 CCGTAGCGGCGGCCGCACTCATACAGCCCGCCCTCGCGGCGGACTTGGCGCAAGACCCGT 2402

Qy      95 ttatccggcaggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcg 154
           | ||     |      | | || ||||| | |||||||| ||| |||||||||| ||||
Db    2403 TCATTACCGATAACGCCCAACGGCAGCACTACGAACCCGGCGGCAAATACCACCTCTTCG 2462

Qy     155 gcagcagggggaac---ttgccgagcgcagcggtcatatcggattgggaaacatacaaa 211
           ||  |  | | |       || |||| |||| |||  | |||    |     |   | |
Db    2463 GCGACCCGCGCGGCAGCGTTTCCGACCGCACCGGCAAAATCAACGTCATCCAAGACTATA 2522

Qy     212 gccatcagttgggcaacctgttcatccagcaggcggccattaaaggaaatatcggctaca 271
           |||  ||| |||||||||||| |||||| |||||    ||| || || |  |||||||||
Db    2523 CCCACCAGATGGGCAACCTGCTCATCCAACAGGCAAACATCAACGGCACAATCGGCTACC 2582

Qy     272 ttgtccgcttttccgatcacgggcacgaagtccattccccccttcgacaaccatgcctcac 331
             ||||||||||||  ||||| |||||||  ||  |||||||||||||||| ||| |
Db    2583 ACACCCGCTTTTCCGGACACGGACACGAAGAACACGCCCCCTTCGACAACCACGCCGCCG 2642

Qy     332 attccgattctgatgaagccggtagtcccgttgacggattcagcctttaccgcatccatt 391
           |   ||      || |||   || |      |  || || || | | | |||||  || | |
Db    2643 ACAGCGCGAGCGAAGAAAAAGGCAACGTTGACGAAGGCTTTACCGTATACCGGCTCAACT 2702

Qy     392 gggacggatacgaacaccatcccgccgacggctatgacgggccacagggcggcggctatc 451
           |||| ||| ||||||| |||||||||| | ||| ||||| ||  ||||||||   || |
Db    2703 GGGAAGGACACGAACATCATCCCGCCGATGCCTACGACGGCCCGAAGGGCGGCAATTACC 2762

Qy     452 ccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaata 511
           ||   || |  ||||  | ||   ||||| ||| || | || |||   ||||  | ||
Db    2763 CCAAACCTACGGGCGCACGAGACGAATACACCTATCACGTCAACGGCACAGCCCGCAGTA 2822

Qy     512 tccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgtcgaccgtttccaca 571
           ||   ||||| | |||||||| |||||||| | | |||||  |   ||||  || | ||
Db    2823 TCAAACTCAATCCGACCGACACCCGCAGCATCCGGCAACGCATATCCGACAATTACAGCA 2882

Qy     572 ataccggtagtatgctgacgcaaggagtaggcgacggattcaaacgcgccacccgataca 631
```

```
Db   2883 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 2942

Qy    632 gccccgagctggacagatcgggcaatgccgccgaagctttcaacggcactgcagatatcg 691
Db   2943 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 3002

Qy    692 tcaaaaacatcatcggcgcggcaggagaaattgtcggc 729
Db   3003 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCGTTGAC 3040


RESULT  28
A96138
LOCUS       A96138                 1497 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 171 from Patent WO9924578.
ACCESSION   A96138
VERSION     A96138.1  GI:6779961
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1497)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 171 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1497
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      475 a    319 c    365 g    338 t
ORIGIN

  Query Match             7.3%;  Score 121.8;  DB 6;  Length 1497;
  Best Local Similarity   66.7%;  Pred. No. 1e-20;
  Matches  174;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
Db    243 CGAACGGACAGGATTTGAAGGTGTTATCGGCTATGAAACCCATTTTTCAGGACACGGACA 302

Qy    297 cgaagtccattccccttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
Db    303 CGAAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 362

Qy    357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
Db    363 CGTAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGC 422

Qy    417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
Db    423 AGACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATAT 482

Qy    477 atacagctacgacataaaagg 497
Db    483 ATACAGCTACCATATCAAAGG 503


RESULT  29
A61829
LOCUS       A61829                 1744 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 9 from Patent WO9711181.
ACCESSION   A61829
VERSION     A61829.1  GI:3715998
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1744)
  AUTHORS   Meyer,T.F., Rudel,T., Scheuerpflug,I., Fischer, Eckhard and
            Maier,J.
  TITLE     NUCLEIC ACID MOLECULES WHICH CODE PROTEINS WHICH MEDIATE THE
            ADHESION OF NEISSERIA CELLS TO HUMAN CELLS
  JOURNAL   Patent: WO 9711181-A 9 27-MAR-1997;
            MAX PLANCK GESELLSCHAFT (DE)
COMMENT     Other publication DE 19534579 970320.
FEATURES             Location/Qualifiers
     source          1. .1744
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
BASE COUNT      560 a    383 c    405 g    396 t
ORIGIN

  Query Match             7.3%;  Score 121.8;  DB 6;  Length 1744;
  Best Local Similarity   66.7%;  Pred. No. 1e-20;
  Matches  174;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
Db    284 CGAACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACA 343

Qy    297 cgaagtccattccccttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
Db    344 CGAAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 403

Qy    357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
Db    404 CGTAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGC 463

Qy    417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
Db    464 AGACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATAT 523

Qy    477 atacagctacgacataaaagg 497
Db    524 ATACAGCTACCATATCAAAGG 544


RESULT  30
A61821
LOCUS       A61821                 3287 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 1 from Patent WO9711181.
ACCESSION   A61821
VERSION     A61821.1  GI:3715995
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 3287)
  AUTHORS   Meyer,T.F., Rudel,T., Scheuerpflug,I., Fischer, Eckhard and
            Maier,J.
  TITLE     NUCLEIC ACID MOLECULES WHICH CODE PROTEINS WHICH MEDIATE THE
            ADHESION OF NEISSERIA CELLS TO HUMAN CELLS
  JOURNAL   Patent: WO 9711181-A 1 27-MAR-1997;
            MAX PLANCK GESELLSCHAFT (DE)
COMMENT     Other publication DE 19534579 970320.
FEATURES             Location/Qualifiers
     source          1. .3287
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
                     /clone_lib="GENOMISCHE BIBLIOTHEK IN PBA"
                     /clone_lib="(B) CLON(E): H1967/PES25"
BASE COUNT     1017 a    740 c    752 g    778 t
ORIGIN

  Query Match             7.3%;  Score 121.8;  DB 6;  Length 3287;
  Best Local Similarity   66.7%;  Pred. No. 1.2e-20;
```

```
  Matches  174;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy     237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
           | | | | | |   || |||| | |||||| ||      ||  ||||| |  ||||| ||
Db    1827 CGAACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACA 1886

Qy     297 cgaagtccattccccctttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
           |||||| ||    || ||||| || ||||  ||| |   |  |||||||    ||| |
Db    1887 CGAAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 1946

Qy     357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
           |  | ||| || | | ||||||   | ||| ||   || |  |||   ||||||||
Db    1947 CGTAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGC 2006

Qy     417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
           ||||| |||||||||||| || |||||||| ||||| |  ||  |||| || ||||||||
Db    2007 AGACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATAT 2066

Qy     477 atacagctacgacataaaagg 497
           |||||||||| | || |||||
Db    2067 ATACAGCTACCATATCAAAGG 2087


RESULT  31
AF142582
LOCUS       AF142582                3294 bp    DNA     linear   BCT 18-MAY-1999
DEFINITION  Neisseria gonorrhoeae maf gene cluster, complete sequence.
ACCESSION   AF142582
VERSION     AF142582.1  GI:4838552
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 3294)
  AUTHORS   Eickernjaeger,S., Meyer,T.F., Fischer,E., Maier,J., Manning,P.A.,
            Rudel,T., Scheuerpflug,I., Schulz,E. and Schwan,E.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-1999) Molekulare Biologie, MPI fuer
            Infektionsbiologie, Monbijoustr. 2, Berlin 10117, Germany
FEATURES             Location/Qualifiers
     source          1. .3294
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
     gene            143. .457
                     /gene="mafI"
     CDS             143. .457
                     /gene="mafI"
                     /codon_start=1
                     /transl_table=11
                     /product="MafI"
                     /protein_id="AAD31037.1"
                     /db_xref="GI:4838553"
                     /translation="MRKSRLSRYKQNKLIELFVAGVTARTAAEPDSIVYTDCYRRYDV
                     LDAGEFSHFRINHSTHFAERQNHINGIGNFWNRAKRHLRKFDGIPKEHFEPYLKECER
                     RF"
     gene            590. .1552
                     /gene="mafA"
     CDS             590. .1552
                     /gene="mafA"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesin MafA"
                     /protein_id="AAD31038.1"
                     /db_xref="GI:4838554"
                     /translation="MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAAS
                     ARAAVKDMDLQALHGRKVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDY
                     TYPRYETTAETTSGGLTGLTTSLSTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNE
                     TLTTNPRDTAFLSHLVQTVFFLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNA
                     ETLKAQTKLEYFAVDRTNKKLLIKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLM
                     VDFSDIRPYGNHTGNSAPSVEADNSHEGYGYSDEAVRQHRQGQP"
     gene            1592. .3121
                     /gene="mafB"
     CDS             1592. .3121
                     /gene="mafB"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesin MafB"
                     /protein_id="AAD31039.1"
                     /db_xref="GI:4838555"
                     /translation="MNLPIQKFMMLFAAAISLLQIPISHANGLDARLRDDMQAKHYEP
                     GGKYHLFGNARGSVKNRVCAVQTFDATAVGPILPITHERTGFEGIIGYETHFSGHGHE
                     VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPADGYDGPQGGGYPEPQGARD
                     IYSYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNW
                     RANRMDDIRGIVQGAVNPFLTGFQGLGVGAITDSAVSPVTYAAARKTLQGIHNLGNLS
                     PEAQLAAASLLQDSAFAVKDGINSARQWADAHPNITATAQTALAVAEAAGTVWGGKKV
                     ELNPTKWDWVKNTGYEKPAARPMQTVDGEMAGKNKPPKPSTQQHSTHSDNNIGLPAPY
                     VKPDTSISPTGTIQDRIRWTKSKFPTEKSLNGHFKAHGKEFGDITIEDYQKMASDLLS
                     KQTSDKILGYQTEHRRVRYDINNNIYVLANPKTFKIKTMFKPNLGKEYYDGEFKKDMG
                     N"
BASE COUNT     1017 a    743 c    755 g    779 t
ORIGIN


  Query Match             7.3%;  Score 121.8;  DB 1;  Length 3294;
  Best Local Similarity   66.7%;  Pred. No. 1.2e-20;
  Matches  174;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy     237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
           | | | | | |   || |||| | |||||| ||      ||  ||||| |  ||||| ||
Db    1834 CGAACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACA 1893

Qy     297 cgaagtccattccccctttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
           |||||| ||    || ||||| || ||||  ||| |   |  |||||||    ||| |
Db    1894 CGAAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 1953

Qy     357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
           |  | ||| || | | ||||||   | ||| ||   || |  |||   ||||||||
Db    1954 CGTAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGC 2013

Qy     417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
           ||||| |||||||||||| || |||||||| ||||| |  ||  |||| || ||||||||
Db    2014 AGACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATAT 2073

Qy     477 atacagctacgacataaaagg 497
           |||||||||| | || |||||
Db    2074 ATACAGCTACCATATCAAAGG 2094


RESULT  32
NMA1Z2491/c
LOCUS       NMA1Z2491               340806 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 1/7.
ACCESSION   AL162752 AL157959
VERSION     AL162752.2  GI:7378778
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 340806)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 340806)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
          sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
          Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT   Notes:
          Details of N. meningitidis sequencing at the Sanger Centre are
          available on the World Wide Web.
          (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .340806
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            complement(38. .358)
                     /gene="NMA0001"
     CDS             complement(38. .358)
                     /gene="NMA0001"
                     /note="NMA0001, unknown, len: 106 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0001"
                     /protein_id="CAB83321.1"
                     /db_xref="GI:7378779"
                     /db_xref="SPTREMBL:Q9JX93"
                     /translation="MSNTQTIRSADFTTSRAWGALDIANMNGTTVRLHWTDQPYKWHV
                     NDGEEVFAVMDGEVDMHYRENGEEHIVRLKSGDIFYAGIGTEHVAHPRGETRILVIEK
                     EGSV"
     gene            complement(414. .2438)
                     /gene="nuoL"
     CDS             complement(414. .2438)
                     /gene="nuoL"
                     /EC_number="1.6.5.3"
                     /note="NMA0002, nuoL, NADH dehydrogenase I chain L, len:
                     674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase
                     I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to
                     entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone
                     (complex I), various chains, and to entry PF00662
                     oxidored_q1_N, NADH-Ubiquinone oxidoreductase (complex I),
                     chain 5 N-terminus"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain L"
                     /protein_id="CAB83322.1"
                     /db_xref="GI:7378780"
                     /db_xref="SWISS-PROT:Q9JX92"
                     /translation="MNDMTLYLIIALVPLAGSLIAGLFGNKIGRAGAHTVTILGVAVS
                     AVLSAYVLWGFLNGSRAKFDENVYTWLTMGGLDFSVGFLVDTMTAMMMVVVTGVSLMV
                     HIYTIGYMHDEKVGYQRFFSYISLFTFSMLMLIMSNNFIQLFFGWEAVGLVSYLLIGF
                     YFKRPSATFANLKAFLINRVGDFGFLLGIGLVLAYFGGSLRYQDVFAYLPNVQNATIQ
                     LFPGVEWSLITVTCLLLFVGAMGKSAQFPLHVWLPDSMEGPTPISALIHAATMVTAGL
                     FMVSRMSPIYEMSSTALSVIMVIGAITALFMGFLGVIQNDIKRVVAYSTLSQLGYMTV
                     ALGASAYSVAMFHVMTHAFFKALLFLAAGSAIIGMHHDQDMRHMGNLKKYMPITWLTM
                     LIGNLSLIGTPFFSGFYSKDSIIEAAKYSTLPGSGFAYFAVLASVFVTAFYAFRQYFM
                     VFHGEEKWRSLPEHHSDGHGEEHHGLGKNDNPHESPLVVTLPLILLAVPSVIIGYIAI
                     EPMLYGDFFKDVIFVNADAHPTMHIMKEEFHGALAMVSHSLHSPVLYLAIAGVLSAWL
                     LYVKLPHLPAKIAQAFRPVYVLFENKYYLDALYFNVFAKGTRALGTFFWKVGDTAIID
                     NGIVNGSARLVGAVAAQVRKVQTGFIYTYAAAMVFGVLVLLGMTFWGLFR"
     misc_feature    1035. .1044
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(1116. .2030)
                     /gene="nuoL"
                     /note="Pfam match to entry PF00361 oxidored_q1,
                     NADH-Ubiquinone/plastoquinone (complex I), various chains,
                     score 351.80, E-value 7.3e-102"
     misc_feature    complement(2061. .2243)
                     /gene="nuoL"
                     /note="Pfam match to entry PF00662 oxidored_q1_N,
                     NADH-Ubiquinone oxidoreductase (complex I), chain 5
                     N-terminus, score 77.20, E-value 3.5e-19"
     gene            complement(2487. .2828)
                     /gene="NMA0003"
     CDS             complement(2487. .2828)
                     /gene="NMA0003"
                     /note="NMA0003, unknown, len: 113 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0003"
                     /protein_id="CAB83323.1"
                     /db_xref="GI:7378781"
                     /db_xref="SPTREMBL:Q9JX91"
                     /translation="MMDKNQLEQEFHKAMLNIYQEALNLPQPYKATRFLQIVNEFGGK
                     EAADKLLSTGEKKTQTGFTELILSGGGVHALKYSMEYLVLQKPWCDLFTEEQLAVARK
                     RLERVGFVSPK"
     RBS             complement(2833. .2837)
     gene            complement(2952. .3527)
                     /gene="NMA0004"
     CDS             complement(2952. .3527)
                     /gene="NMA0004"
                     /note="NMA0004, unknown, len: 191 aa; similar to
                     hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa),
                     fasta scores; E(): 0; 80.5% identity in 190 aa overlap,
                     and to proteins involved in cell cycle e.g. TR:O54679
                     (EMBL:AF036487) Lactococcus lactis Plasmid pNZ4000
                     putative mobilization protein (200 aa), fasta scores; E():
                     0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell
                     filamentation protein FIC (200 aa), fasta scores; E():
                     0.093, 29.4% identity in 126 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0004"
                     /protein_id="CAB83324.1"
                     /db_xref="GI:7378782"
                     /db_xref="SPTREMBL:Q9JQR9"
                     /translation="MPSENPIGKTMKSIDEQSLHNARRLFESGDIDRIEVGTTAGLQQ
                     IHRYLFGGLYDFAGQIREDNISKGGFRFANAMYLKEALVKIEQMPERTFEEIIAKYVE
                     MNIAHPFLEGNGRSTRIWLDLVLKKNLKKVVNWQNVSKTLYLQAMERSPVNDLELRFL
                     LKDNLTDDVDNREIIFKGIEQSYYYEGYEKG"
     misc_feature    complement(3516. .3525)
                     /gene="NMA0004"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(3561. .3866)
                     /gene="nuoK"
     CDS             complement(3561. .3866)
                     /gene="nuoK"
                     /EC_number="1.6.5.3"
                     /note="NMA0005, nuoK, NADH dehydrogenase I chain K, len:
                     101 aa; simlar to many e.g. NUOK_RHOCA NADH dehydrogenase
                     I chain K (EC 1.6.5.3) (102 aa), fasta scores; E():
                     1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam
                     match to entry PF00420 oxidored_q2,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain K"
                     /protein_id="CAB83325.1"
                     /db_xref="GI:7378783"
                     /db_xref="SPTREMBL:Q9JQU9"
                     /translation="MITLTHYLVLGALLFGISAMGIFMNRKNVLVLLMSIELMLLAVN
                     FNFIAFSQHLGDTAGQIFVFFVLTVAAAESAIGLAIMVLVYRNRQTINVADLDELKG"
     misc_feature    complement(3564. .3863)
                     /gene="nuoK"
                     /note="Pfam match to entry PF00420 oxidored_q2,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
                     score 141.50, E-value 1.5e-38"
     gene            complement(3863. .4534)
                     /gene="nuoJ"
     CDS             complement(3863. .4534)
                     /gene="nuoJ"
                     /EC_number="1.6.5.3"
                     /note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
                     223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
                     I chain J (EC 1.6.5.3) (202 aa), fatsa scores; E():
                     9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
                     match to entry PF00499 oxidored_q3,
```

```
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain J"
                     /protein_id="CAB83326.1"
                     /db_xref="GI:7378784"
                     /db_xref="SPTREMBL:Q9JX90"
                     /translation="MTFQLILFYIFAVIILYGAIKTVTAKNPVHAALHLVLTFCVSAM
                     LWMLMQAEFLGVTLVVVYVGAVMVLFLFVVMMLNIDIEEMRAGFWRHAPVAGVVGTLL
                     AVALILILVNPKTDLAAFGLMKDIPADYNNIRDLGSRIYTDYLLPFELAAVLLLGMV
                     AAIALVHRKTVNPKRMDPADQVKVRADQGRMRLVKMEAVKPQTESAEESEVSDDLKPK
                     EEGKA"
     RBS             complement(3874. .3879)
                     /gene="nuoJ"
     misc_feature    complement(4034. .4525)
                     /gene="nuoJ"
                     /note="Pfam match to entry PF00499 oxidored_q3,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
                     score 77.90, E-value 2e-19"
     gene            complement(4570. .5277)
                     /gene="NMA0007"
     CDS             complement(4570. .5277)
                     /gene="NMA0007"
                     /note="NMA0007, unknown, len: 235 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0007"
                     /protein_id="CAB83327.1"
                     /db_xref="GI:7378785"
                     /db_xref="SPTREMBL:Q9JX89"

  Query Match              6.6%;  Score 110.2;  DB 1;  Length 340806;
  Best Local Similarity   62.5%;  Pred. No. 3.5e-17;
  Matches  172;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy     237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
           | | | | | |   || |||| | |||||| ||      ||  ||||| |  || || ||
Db  305687 CGAACGGACAGGATTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACATGGACA 305628

Qy     297 cgaagtccattccccctttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
            ||||| ||    || ||||| || ||||  ||| |   |  |||||||     |||  |
Db  305627 TGAAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 305568

Qy     357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
             |  |  || || | | ||||||   | ||| ||   || |  |||  |||||| |
Db  305567 CGTAGACGGTGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATCCATCCGGA 305508

Qy     417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
            || || |||||||||| || ||| |||  |||||  | |||  ||| || ||||||||
Db  305507 GGATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATAT 305448

Qy     477 atacagctacgacataaaaggcgttgcccaaaata 511
           |||||||||| |  | |||||   | |   ||| |
Db  305447 ATACAGCTACTATGTCAAAGGAACTTCAACAAAAA 305413


RESULT  33
A96134
LOCUS       A96134                   1449 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 167 from Patent WO9924578.
ACCESSION   A96134
VERSION     A96134.1  GI:6779959
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1449)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 167 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1449
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      445 a    310 c    335 g    340 t     19 others
ORIGIN


  Query Match              6.4%;  Score 107.2;  DB 6;  Length 1449;
  Best Local Similarity   61.5%;  Pred. No. 6.5e-17;
  Matches  169;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy     237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
           | | | | | |   || |||| | |||||| ||      ||  ||||| |  || || ||
Db     243 CGAACGGACAGGATTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACATGGACA 302

Qy     297 cgaagtccattccccctttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
            ||||| ||    || ||||| || ||||  ||| |   |  |||||||     |||  |
Db     303 TGAAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 362

Qy     357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
             |  |  || || | | ||||||   | ||| ||   || |  |||  |||||| |
Db     363 CGTAGACGGTGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATCCATCCGGA 422

Qy     417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
            || || |||||||||| || ||| |||  |||||  | |||  ||| || ||||||||
Db     423 GGATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATAT 482

Qy     477 atacagctacgacataaaaggcgttgcccaaaata 511
           |||||  ||  |  | |||||  | |   ||| |
Db     483 ATACANNTANTATGTCAAAGGAACTTCAACAAAAA 517


RESULT  34
A96132
LOCUS       A96132                   1452 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 165 from Patent WO9924578.
ACCESSION   A96132
VERSION     A96132.1  GI:6779958
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1452)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 165 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1452
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      466 a    298 c    332 g    356 t
ORIGIN


  Query Match              6.4%;  Score 107;  DB 6;  Length 1452;
  Best Local Similarity   61.8%;  Pred. No. 7.4e-17;
  Matches  170;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
           | | | | | |   || ||||   |||||| ||      ||  ||||| |  ||||| ||
Db     243 CGAACGGACAGGGTTTGAAGGTGTTATCGGTTATGAAACCCATTTTTCAGGGCACGGACA 302

Qy     297 cgaagtccattccccctttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
            ||||| ||    || |||||  | ||||  ||| |   |  |||||||     |||  |
Db     303 TGAAGTACACAGTCCGTTCGATCATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 362

Qy     357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
           |   |  | ||| || |   ||||||   | ||| |    || |  |||  |||||| |
Db     363 TGTAGACGGCGGTTTTACTGTTTACCAACTTCATCGAACAGGGTCGGAAATCCATCCGGA 422

Qy     417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
```

```
          || || |||||||||||| || ||| |||  |||||  | |||  ||| || ||||||||
Db    423 GGATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATAT 482

Qy    477 atacagctacgacataaaaggcgttgcccaaaata 511
          |||||||||  |  | |||||   | |   ||| |
Db    483 ATACAGCTATTATGTCAAAGGAACTTCAACAAAAA 517


RESULT  35
AE002559
LOCUS       AE002559               10869 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 201 of 206
            of the complete genome.
ACCESSION   AE002559 AE002098
VERSION     AE002559.2  GI:7413482
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10869)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10869)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Apr 4, 2000 this sequence version replaced gi:7227357.
FEATURES             Location/Qualifiers
     source          1. .10869
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            150. .524
                     /gene="NMB2095"
     CDS             150. .524
                     /gene="NMB2095"
                     /note="similar to PID:1817528 percent identity: 53.99;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesin complex protein, putative"
                     /protein_id="AAF42412.1"
                     /db_xref="GI:7227358"
                     /translation="MKLLTTAILSSAIALSSMAAAAGTDNPTVAKKTVSYVCQQGKKV
                     KVTYGFNKQGLTTYASAVINGKRVQMPVNLDKSDNVETFYGKEGGYVLGTGVMDGKSY
                     RKQPIMITAPDNQIVFKDCSPR"
     gene            complement(712. .2178)
                     /gene="NMB2096"
     CDS             complement(712. .2178)
                     /gene="NMB2096"
                     /note="similar to SP:P33940 PID:424051 GB:U00096 percent
                     identity: 69.73; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="malate:quinone oxidoreductase"
                     /protein_id="AAF42413.1"
                     /db_xref="GI:7227359"
                     /translation="MAEATDVVLVGGGIMSATLGVLLKELEPSWEITLIERLEDVALE
                     SSNAWNNAGTGHSALCELNYAPLGANGIIDPARALNIAEQFHVSRQFWATLVAEGKLE
                     DNSFINAVPHMSLVMNEDHCSYLQKRYDAFKTQKLFENMEFSTDRNKISDWAPLMMRG
                     RDENQPVAANYSAEGTDVDFGRLTRQMVKYLQGKGVKTEFNRHVEDIKRESDGAWVLK
                     TADTRNPDGQLTLRTRFLFLGAGGGALTLLQKSGIPEGKGYGGFPVSGLFFRNSNPET
                     AEQHNAKVYGQASVGAPPMSVPHLDTRNVDGKRHLMFGPYAGFRSNFLKQGSLMDLPL
                     SIHMDNLYPMLCAGWANMPLTKYLLGELRKTKEERFASLLEYYPEANPDDWELITAGQ
                     RVQIIKKDSEKGGVLQFGTEIVAHADGSLAALLGASPGASTAVPLMIRLMHQCFPERA
                     PSWEDRLKELVPGYGIKLNENPERADEIIAYTAKVLDI"
gene            complement(2230. .2319)
                /gene="NMB2097"
CDS             complement(2230. .2319)
                /gene="NMB2097"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF42414.1"
                /db_xref="GI:7227360"
                /translation="MLATHSNPETAYSDTHHCRFSLLLRFRRL"
gene            complement(2715. .3305)
                /gene="NMB2098"
CDS             complement(2715. .3305)
                /gene="NMB2098"
                /note="conserved hypothetical protein; identified by
                Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF42415.1"
                /db_xref="GI:7227361"
                /translation="MSEKIILPVLSLGGVRLEPLDVHHETGLREAVCDGEVWKLGVTS
                APHPDRVADYIGTALATRLAFAVVDEEAGRVVGTTAYYHFEPQIPRLDIGFTWYAASA
                RRTRINTCCKIMLLDSAFDVLACRCVGWRTDILNLASQRAIERLGAEKDGVLRMHMLR
                KDGSVRDTVVYSMLREDWCKNREILTGRLAGYGVQV"
gene            complement(3366. .3947)
                /gene="NMB2099"
CDS             complement(3366. .3947)
                /gene="NMB2099"
                /note="conserved hypothetical protein; identified by
                Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF42416.1"
                /db_xref="GI:7227362"
                /translation="MRNEEKRALRRELRGRRSQMGRDVRAAATVKINHLLKRYIKKGR
                KIGVYWPMGKELRLDGFVRAAQKRGAELYLPYIEPRSRRMWFTPYPADGVKQERKRGR
                AKLHVPQFAGRKKRVHDLNLLLVPVVGMDRLGYRLGQAGGYYDATLSAMKYRLQAKTV
                GVGFACQLVDRLPVEAHDRSLDGFVSEAGILCF"
gene            4025. .4258
                /gene="NMB2100"
CDS             4025. .4258
                /gene="NMB2100"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF42417.1"
                /db_xref="GI:7227363"
                /translation="MWHIVAIGYLFVAVMYSAAQPSIARALIYLVFWAVLPTVFTVFT
                ITVRRRNHLMRQQEQAESEQQRAQRQKDSGTKP"
gene            4431. .5159
```

```
                     /gene="NMB2101"
     CDS             4431. .5159
                     /gene="NMB2101"
                     /note="similar to SP:P34831 GB:X53651 PID:296030 percent
                     identity: 78.02; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="30S ribosomal protein S2"
                     /protein_id="AAF42418.1"
                     /db_xref="GI:7227364"
                     /translation="MSQITMRQMIEAGVHFGHQTRFWNPKMAQYIFGARNKIHIVNLE
                     KTLPMFQDAQEAVRRLVANKGTVLFVGTKRQARDIIREEATRAGMPFVDYRWLGGMLT
                     NYKTVKQSIKRLEEKTAALENAAESGFSKKEILEMQRDVEKLERSLGGIKNMKGLPDA
                     IFVIDTGYQKGTLVEAEKLGIPVIAVVDTNNSPDGVKYVIPGNDDSAKAIRLYCRGIA
                     DAVLEGKNQALQETVAAAQEAAAE"
     gene            5288. .6142
                     /gene="NMB2102"
     CDS             5288. .6142
                     /gene="NMB2102"
                     /note="similar to GB:D26562 SP:P02997 GB:D13334 GB:V00343
                     PID:1208943 percent identity: 69.01; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="elongation factor TS (EF-TS)"
                     /protein_id="AAF42419.1"
                     /db_xref="GI:7227365"
                     /translation="MAEITAKMVADLRAATGLGMMECKKALVEAEGNFDKAEEILRIK
                     SGAKAGKLAGRTAAEGVLAYAINGNVGALVEVNCETDFVAKDAGFVEFANFVAKTAAE
                     KKPASVEELSELVEAERKAIIAKLGENMSVRRFQVIDTANQLVAYIHGALATEGVLVE
                     YKGSEDVARKIGMHIVAAKPQCVSEAEVDAETVEKERHIYTEQAIASGKPADIAAKMV
                     EGRIRKFLAEITLNGQAFVMNPDQTVAQFSKENGTEVISFVRYKVGDGIEKKAVDYAA
                     EVAAAAKV"
     gene            6359. .7078
                     /gene="NMB2103"
     CDS             6359. .7078
                     /gene="NMB2103"
                     /note="similar to GB:D26562 SP:P29464 GB:D13334 GB:X78809
                     PID:1208944 percent identity: 73.16; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="uridylate kinase"
                     /protein_id="AAF42420.1"
                     /db_xref="GI:7227366"
                     /translation="MTQQIKYKRVLLKLSGESLMGSDPFGINHDTIVQTVGEIAEVVK
                     MGVQVGIVVGGGNIFRGVSAQAGSMDRATADYMGMMATVMNALALKDAFETLGIKARV
                     QSALSMQQIAETYARPKAIQYLEEGKVVIFAAGTGNPFFTTDTAAALRGAEMNCDVML
                     KATNVDGVYTADPKKDPSATRYETITFDEALLKNLKVMDATAFALCRERKLNIVVFGI
                     AKEGSLKRVITGEDEGTLVHC"
     gene            7262. .8225
                     /gene="NMB2104"
                     /note="similar to GP:4838554 percent identity: 99.38;
                     identified by sequence similarity; putative"
     gene            8265. .9716
                     /gene="NMB2105"
     CDS             8265. .9716
                     /gene="NMB2105"
                     /note="similar to GP:4838555 percent identity: 0.00;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="mafB protein"

  Query Match              6.4%;  Score 107;  DB 1;  Length 10869;
  Best Local Similarity   61.8%;  Pred. No. 1.1e-16;
  Matches  170;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
           | | | | | |   || ||||   |||||| ||      ||  ||||| |  ||||| ||
Db    8507 CGAACGGACAGGGTTTGAAGGTGTTATCGGTTATGAAACCCATTTTTCAGGGCACGGACA 8566

Qy     297 cgaagtccattcccccttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
           ||||| ||    || |||||  | ||||  ||| |   |  |||||||     |||  |
Db    8567 TGAAGTACACAGTCCGTTCGATCATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 8626

Qy     357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
           |   | | ||| || |   ||||||   | ||| |    || |  |||  ||||||  |
Db    8627 TGTAGACGGCGGTTTTACTGTTTACCAACTTCATCGAACAGGGTCGGAAATCCATCCGGA 8686

Qy     417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
           || || ||||||||||| || ||| |||  |||||  | |||  ||| || ||||||||
Db    8687 GGATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATAT 8746

Qy     477 atacagctacgacataaaaggcgttgcccaaaata 511
           |||||||||  |  | |||||   | |   ||| |
Db    8747 ATACAGCTATTATGTCAAAGGAACTTCAACAAAAA 8781


RESULT  36
AX044035
LOCUS       AX044035              172325 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 114 from Patent WO0066791.
ACCESSION   AX044035
VERSION     AX044035.1  GI:11342919
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 172325)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 114 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .172325
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.-seq 1: 1 to
                     349980 349980 bases-seq 108: 300001 to 649980 349980
                     bases-seq 109: 600001 to 949980 349980 bases-seq 110:
                     900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
                     349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
                     113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    43072 a  47583 c  41465 g  40205 t
ORIGIN


  Query Match              6.4%;  Score 107;  DB 6;  Length 172325;
  Best Local Similarity   61.8%;  Pred. No. 2.1e-16;
  Matches  170;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
           | | | | | |   || ||||   |||||| ||      ||  ||||| |  ||||| ||
Db 122937 CGAACGGACAGGGTTTGAAGGTGTTATCGGTTATGAAACCCATTTTTCAGGGCACGGACA 122996

Qy     297 cgaagtccattcccccttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
           ||||| ||    || |||||  | ||||  ||| |   |  |||||||     |||  |
Db 122997 TGAAGTACACAGTCCGTTCGATCATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 123056

Qy     357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
           |   | | ||| || |   ||||||   | ||| |    || |  |||  ||||||  |
Db 123057 TGTAGACGGCGGTTTTACTGTTTACCAACTTCATCGAACAGGGTCGGAAATCCATCCGGA 123116

Qy     417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
           || || ||||||||||| || ||| |||  |||||  | |||  ||| || ||||||||
Db 123117 GGATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATAT 123176

Qy     477 atacagctacgacataaaaggcgttgcccaaaata 511
           |||||||||  |  | |||||   | |   ||| |
Db 123177 ATACAGCTATTATGTCAAAGGAACTTCAACAAAAA 123211
```

```
RESULT  37
A96130
LOCUS       A96130                   375 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 163 from Patent WO9924578.
ACCESSION   A96130
VERSION     A96130.1  GI:6779957
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 163 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      105 a     89 c     90 g     90 t      1 others
ORIGIN
```

Query Match 5.9%; Score 98; DB 6; Length 375;
Best Local Similarity 59.6%; Pred. No. 1.2e-14;
Matches 164; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy    237 ccagcaggcggccattaaaggaaatatcggctacattgtccgcttttccgatcacgggca 296
          | | | | | |   || |||   |||||| ||      ||  ||||| |  ||||| ||
Db     27 CGAACGGACAGGGTTTGAAGGTGTTATCGGTTATGAAACCCATTTTTCAGGGCACGGACA 86

Qy    297 cgaagtccattccccccttcgacaaccatgcctcacattccgattctgatgaagccggtag 356
           ||||| ||    || |||||  | ||||  ||| |   |  |||||||     |||  |
Db     87 TGAAGTACACAGTCCGTTCGATCATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGG 146

Qy    357 tcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgc 416
          |   | | ||| || |   ||||||   | ||| |     | |  |||  |||||| |
Db    147 TGTAGACGGCGGTTTTACTGTTTACCAACTTCATCGAACATGGTCGGAAATCCATCCGGA 206

Qy    417 cgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatat 476
           || |  ||||||||||| || |  |     |||||  | |||  ||| || |||||||||
Db    207 GGATGAATATGACGGGCCGCAAGCAGCGNATTATCCGCCCCCCGGAGGAGCAAGGGATAT 266

Qy    477 atacagctacgacataaaaggcgttgcccaaaata 511
          |||||||||  |  | |||||   | |   ||| |
Db    267 ATACAGCTATTATGTCAAAGGAACTTCAACAAAAA 301
```

```
RESULT  38
AF429315
LOCUS       AF429315                 125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN
```

Query Match 3.1%; Score 51.8; DB 9; Length 125020;
Best Local Similarity 10.6%; Pred. No. 0.053;
Matches 103; Conservative 414; Mismatches 445; Indels 8; Gaps 3;

```
Qy    262 atcggctacattgtccgcttttccgatcacgggcacgaagtccattccccccttcgacaac 321
          :| : |  :::::   |: :  ||    :::|  ::::    :    :    :: :: ::
Db  16923 MTWSCCMSMMYKSSWWGSSWGMCCWGRRRSKGWKWYSRGMSRSKSMRYTGGSKMMRSSMM 16982

Qy    322 catgcctcacattccgattctgatgaagccggtagtcccgttgacggattcagcctttac 381
          |  :|: ::|   | :   : :   : :::   :::||  :  ::: :| :|:::    :
Db  16983 CTSSCYASMCCMCCWSCCCMRSCCCMCRSYCCCMRYCCACKCYMSSYWTMSASYMSRRSY 17042

Qy    382 cgcatccattgggacggatacgaacaccatcccgccgacggctatgacgggccacagggc 441
          :    |:    ::::: :  :: |  ::     |  :  :: ::::: : :
Db  17043 WKRSMKCWSRMCSRSRSSKCKSRGCGGSMGKGKGGKSYGRKTKRKSRGKMKGAKWYMYYR 17102

Qy    442 ggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgtt 501
          :::::::::::: |::|:: :  | :   : :|    :| ::   :
Db  17103 RSMRKMKMYSSKGMYCMYCWCWGRRGCYCSCMTSRSAMCCSYCYAKCKSMCYSCYYGMSM 17162

Qy    502 gcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgtcgac 561
          : :: |  :: :::: :|:: :|   :  |        | ||       |     |  |
Db  17163 KGYYYACSYRGSMSSKYCMRGSTYSTSCGCCCTTTTTCCCCCNANTGGGGAAGCTTTTNC 17222

Qy    562 cgtttccacaataccggtagtatgctgacgcaaggagtaggcgacggattcaaacgcgcc 621
          ::|:::   :  |   :    :   :::|:||   |  ::::  ::  : : ::: :|
Db  17223 NKKTYSYYRKRNGCAMCKYNNYYNNSWRSSCRAGMSCTKYKSSMTMSMASYCWCMSMYYC 17282

Qy    622 acccgatacagccccgagctggacagatcgggcaatgccgccgaagctttcaacggcact 681
          : : | : :: :::   ::::: :: :::    :: |: ::: :::   :  : ::
Db  17283 SMRSMASRGMSWSSYMKMMKSSWRMSCYMKCCWCMKMCYCMMRSMRSRGMSYYMYASWKS 17342

Qy    682 gcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgcaggcgatgcc 741
          :::| : |  :  :   : :  ::  :::::: |    : : :  |: :  : :: ::::
Db  17343 SSRGCYCTRCYWCMSSKSCYKSYYMMMRSKRMKGMKWSRCWGSMWGSASRSSCYKCYKSM 17402

Qy    742 gtgcagggtataagcgaaggctcaaacattgctgttatg--cacggcttgggtctgcttt 799
          : ::::::    ||| :::| ::: :  :    ::: ::  ::::::::|:| :  | :
```

```
Db  17403 RCSMMSSKCYRCAGCMMMKGGYMRYMRCWSMKRRWWGKSAMYRMRRMWKGRGAMMMCMKC 17462

Qy    800 ccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactcaaagact 859
          :::::: ::::      :  :| ::    ::    :  : : : | | : :::::   :|:
Db  17463 YSRMSRCMMMMKSYCASCRSCAMMSGTYMKCASYMCMSSYWCTCWCTSYMSYYRCTCWCK 17522

Qy    860 atgccgcagcagccatccgcgattgggc----agtccaaaaccccaatgccgcacaaggc 915
           ::::   ::::::: ::: |: :: ::    :| :: :::||::::  ::::  :: ::
Db  17523 GWSSYTKSKSWSSSMSSYKKGRRKSYSMCCTSRGAMSCWRRCCYMRGASSMRAGSMSRRR 17582

Qy    916 atagaagccgtcagcaatatctttacggcagtcatccccgtcaaagggattggagctgtt 975
          |: : :|  :: :     : :::: :   :: ::|::||  |::::: :::::::  |
Db  17583 AKGRSWGGRSKWWMTGGMWRSKYYYYCTGRRMMMTYMCC--CWRRRSYMYYRSMSAMGMR 17640

Qy    976 cggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatgggc 1035
          :: : :::  : ::   :: ::: :|  :   :: :|::  ::| ::  :  :  :|:
Db  17641 KSSWSGMRMMGSASSRRCKSASRSSWCSRRMKGMRGSCWSSKMWGGSRSRSASSKCKGSR 17700

Qy   1036 gagatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggcatac 1095
          |::: ::   :   : : ::: : :: |: |::   :::   |  ::  ::  :  :
Db  17701 GMRRRSKRSSKYRKRGRGKKRSMTKSKGSKGKSKCWKRSGSMTSSCYYYSASSCMWMMSS 17760

Qy   1096 gccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgttac 1155
          ::|:   ::: : | ||:  :: | ::   :  :::::|      :  | :  :  ::::
Db  17761 KSCMCCCMMMKRRCACCYSSMSCTSMYRCCWGGMKSYSTCGTTCKKCTGYKKSRTWYWWM 17820

Qy   1156 ggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggaaagaatgtgaaactg 1215
           :      ::: | :::|:::| :::|  ::|   |:|:  ::|| |  |    | | ||
Db  17821 YSWTSKMYMWRAACMMYCYMSTRRMMCMMSMCARKGWCMCTMKGAGATCACAGTAGAGTG 17880

Qy   1216 gcaaacaaac 1225
            | | ||||
Db  17881 AGATAGAAAC 17890


RESULT 39
AE002514
LOCUS       AE002514               10947 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 156 of 206
            of the complete genome.
ACCESSION   AE002514 AE002098
VERSION     AE002514.1  GI:7226886
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10947)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10947)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10947
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            815. .1634
                     /gene="NMB1636"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     PID:2367651 percent identity: 100.00; identified by
                     sequence similarity; putative;opacity protein, authentic
                     frameshift"
     gene            complement(1835. .2392)
                     /gene="NMB1637"
     CDS             complement(1835. .2392)
                     /gene="NMB1637"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF41986.1"
                     /db_xref="GI:7226887"
                     /translation="MDALKLLTNRRSSKKLKHPAPDAAELEQIFQAATQVPDHGNMRP
                     FRFTVIQGEVGLQRFRDVLKQTVAELNFGDDAMKKAEKVGNMAPMVIGVTFAPNRDVP
                     KPKPEWEQMLTAGCAAYALQLAATAQGFDNVWITGMWVNSPLLREAFGCADKDKIIGL
                     MMVGTPTEEVHKPKNTDLEAFVSHW"
     gene            complement(2468. .4102)
                     /gene="NMB1638"
     CDS             complement(2468. .4102)
                     /gene="NMB1638"
                     /note="similar to GB:U14003 SP:P30845 GB:D14055 PID:216537
                     PID:536958 percent identity: 64.40; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="YhbX/YhjW/YijP/YjdB family protein"
                     /protein_id="AAF41987.1"
                     /db_xref="GI:7226888"
                     /translation="MIKPNLRPKLGSSALIAFLSLYSSLVLNYAFFAKVVELHPFNGT
                     GADIFLYTMPVVLFFLSNFVFHVIALPFVHKVLIPLILVISAAVSYQEIFFNIYFNKS
                     MLNNVLQTTAAESARLITPGYVLWIVCLGVLPALAYIAVKVKYRVWYKELLTRLVLAA
                     VSFLCALGIAMLQYQDYASFFRNNKSVTHLIVPSNFIGAGVSKYKDWKRSNIPYTQLD
                     MAVVQNRPAGSLRRFVVLVVGETTRAANWGLNGYSRQTTPLLAARGDEIVNFPQVRSC
                     GTSTAHSLPCMFSTFDRTDYDEIKAEHQDNLLDIVQRAGVEVTWLENDSGCKGVCGKV
                     PNTDVTSLNLPEYCRNGECLDNILLTKFDEVLNKNDKDAVLILHTIGSHGPTYYERYT
                     EAERKFTPTCDTNEINKCTRATLVNTYDNTVLYVDQFIDKVIRKLENRDDLESVVHYV
                     SDHGESLGENGMYLHAAPYAIAPSGQTHIPMVMWFSKAFRQHGGIDFQCLKQKAAENE
                     YSHDHYFSTVLGLMDISNSQTYRKEMDILAACRRPR"
     gene            complement(4176. .4505)
                     /gene="NMB1639"
     CDS             complement(4176. .4505)
                     /gene="NMB1639"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41988.1"
                     /db_xref="GI:7226889"
                     /translation="MSRCCLKRLQTASAVSAFSGGVLEVVRLCCRRLSVLFNPSVMLF
                     VPDCPVGGAAGFFQYEMLCPFFWQGRLQTGSNLAYDVFMSAGRLNGGRNPPHSRRFLA
                     LLCSVAL"
     gene            complement(4505. .5611)
                     /gene="NMB1640"
     CDS             complement(4505. .5611)
                     /gene="NMB1640"
                     /note="similar to PID:2253231 PID:2253236 PID:2253251
                     percent identity: 100.00; identified by sequence
                     similarity; putative"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoserine aminotransferase"
                     /protein_id="AAF41989.1"
                     /db_xref="GI:7226890"
                     /translation="MSLYPIYNFSAGPAVLPEAVLETARQEMLDYNGTGFPVMAMSHR
                     SEMFLSILHHAEQDLRQLLKVPDNYKILFLQGGATTQFNMAAMNLAHGFRTADAVVTG
                     NWSRIAYEQMSRLTDTEIRLAAHGGEQFDYLDLPPVETWDVAPDSAFVHFAVNETVNG
                     LQYREVPCLSEGMPPLVCDMSSEILSREFDVADYGLIYAGAQKNIGPAGVTVVIVRED
                     LLERCPNDIPDVFNYRSHINRDGMYNTPSTYAIYMSGLVFRWLQAQGGVKKIEAVNRL
                     KAQTLYETIDGSDGFYINRIRPNARSKMNVVFQTGDEELDRRFVLEAELQGLCLLKGY
                     KTVGGMRASIYNAMPLEGVRALADFMRDFQRRYG"
gene                 5958. .6407
                     /gene="NMB1641"
CDS                  5958. .6407
                     /gene="NMB1641"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF41990.1"
                     /db_xref="GI:7226891"
                     /translation="MYIGSSMDIQTILEKTLPGLGYELVDFELTAQGTLRVFIDKESG
                     ITVEDCATVSNHLSRVFMVEDIDYKNLEISSPGLDRPLKKAADFVRFAGQNAKIKTRL
                     PIDGQKNFIGKIEGCENDTVTVSFDGKTVQIELGNIDKARLRPEFKF"
gene                 6435. .7937
                     /gene="NMB1642"
CDS                  6435. .7937
                     /gene="NMB1642"
                     /note="similar to GB:L42023 SP:P43915 PID:1007236
                     PID:1221412 PID:1205522 percent identity: 70.34;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="N utilization substance protein A"
                     /protein_id="AAF41991.1"
                     /db_xref="GI:7226892"
                     /translation="MSREMLQLAEALASEKNVDAEVVFQALEFALSTAAKKKADREHM
                     DVRVQINRDTGEYQTFRRWLIVADEDYTYPDVEKTIEEIQEEIPGTTIQIGEYYEEQL
                     PNEGFGRQAAQTAKQIILQRIRDAEREQNLNEFLAVKEDIVSGTVKRVERHGIIVEVV
                     AGKLDALIPRDQMIPRENFRSGDRIRALFLRVEEIGNTGRKQVILSRTSGDFLVKLYA
                     NEVPEIADGMLEIRAVARDPGQRAKVAVKANDQRIDPQGTCIGVRGSRVNAVSNELSG
                     ERIDVVLWSPEPAQFVMSALSPAEVSRIVIDEDKHAVDVIVAEDQLALAIGRGGQNVR
                     LASDLTGWQLNIMTSAEADERNAAEDAAIRRLFMDHLNVDEETADVLVQEGFATLEEV
                     AYVPAAELLAIEGFDEEIVDMLRNRARDAILTMAIAAEEKLGEVSDDMRNLEGIDADM
                     LRSLAEAGITTRDDLAELAVDELIEITGVNEETAKAVILTAREHWFTEDK"
gene                 7949. .10837
                     /gene="NMB1643"
CDS                  7949. .10837
                     /gene="NMB1643"
                     /note="similar to GP:2597869 percent identity: 73.73;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="translation initiation factor IF-2"
                     /protein_id="AAF41992.1"
                     /db_xref="GI:7226893"
                     /translation="MSNTTVEQFAAELKRPVEDLLKQLKEAGVSKNSGSDSLTLDDKQ
                     LLNAYLTKKNGSNSSTISIRRTKTEVSTVDGVKVETRKRGRTVKIPSAEELAAQVKAA
                     QTQAAPVRPEQTAEDAAKARAEAAARAEARAKAEAEAAKLKAAKAGNKAKPAAQKPTE
                     AKAETAPVAAETKPAEESKAEKAQADKMPSEKPAEPKEKAAKPKHERNGKGKDAKKPA
                     KPAAPAVPQPVVSAEEQAQRDEEARRAAALRAHQEALLKEKQERQARREAMKQQAEQQ
                     AKAAQEAKTGRQRPAKPAEKPQAAAPAVENKPVNPAKAKKEDRRNRDDEGQGRNAKGK
                     GGKGGRDRNNARNGDDERVRGGKKGKKLKLEPNQHAFQAPTEPVVHEVLVPETITVAD
                     LAHKMAVKGVEVVKALMKMGMMVTINQSIDQDTALIVVEELGHIGKPAAADDPEAFLD
                     EGAEAVEAEALPRPPVVTVMGHVDHGKTSLLDYIRRTKVVQGEAGGITQHIGAYHVET
                     PRGVITFLDTPGHEAFTAMRARGAKATDIVILVVAADDGVMPQTIEAIAHAKAAGVPM
                     VVAVNKIDKEAANPERIRQELTAHEVVPDEWGGDVQFIDVSAKKGLNIDALLEAVLLE
                     AEVLELTAPVDAPAKGIIVEARLDKGRGAVATLLVQSGTLKKGDMLLAGTAFGKIRAM
                     VDENGKSITEAGPSIPVEILGLSDVPNAGEDAMVLADEKKAREIALFRQGKYRDVRLA
                     KQQAAKLENMFNNMGETQAQSLSVIIKADVQGSYEALAGSLKKLSTDEVKVNVLHSGV
                     GGITESDVNLAIASGAFIIGFNVRADASSRKLAENENVEIRYYNIIYDAINDVKAAMS
                     GMLSPEEKEQVTGTVEIRQVISVSKVGNIAGCMVTDGVVKRDSHVRLIRNNVVIHTGE
                     LASLKRYKDDVKEVRMGFECGLMLKGYNEIMEGDQLECFDIVEVARSL"
BASE COUNT      2921 a    3105 c    2733 g    2188 t
ORIGIN


  Query Match             2.7%;  Score 45.4;  DB 1;  Length 10947;
  Best Local Similarity   55.4%;  Pred. No. 1.5;
  Matches  108;  Conservative    0;  Mismatches   86;  Indels    1;  Gaps    1;

Qy     538 agcaccggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaagga 597
           ||||| |||   || | |||| ||||   |||  | |  || ||   ||  || || |||
Db    6509 AGCACTGGAATTCGCCCTGTCTACCGCCGCCAAGAAAAAGGCAGACCGCGAACACATGGA 6568

Qy     598 -gtaggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaa 656
            ||  ||| |  | |||| |||| |||| |  |   ||       |   | | | |  |
Db    6569 CGTGCGCGTCCAAATCAACCGCGACACCGGCGAATACCAAACCTTCCGCCGCTGGCTGAT 6628

Qy     657 tgccgccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcagg 716
           || |||||| |    | |   |  | | ||| |||  |||| ||||   |       ||
Db    6629 TGTCGCCGATGAAGACTATACCTATCCCGATGTCGAAAAAACCATCGAGGAAATCCAAGA 6688

Qy     717 agaaattgtcggcgc 731
           ||||||  |||| |
Db    6689 GGAAATTCCCGGCAC 6703


RESULT  40
NMA6Z2491
LOCUS       NMA6Z2491             326301 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 6/7.
ACCESSION   AL162757 AL157959
VERSION     AL162757.2  GI:7380371
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 326301)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 326301)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .326301
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     repeat_unit     196. .311
                     /note="REP 2; hmmfs hit to HMM REP 2 (1 - 109), score:
                     87.43"
     repeat_unit     434. .453
                     /note=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
                     /label=dRS3
     repeat_unit     498. .560
```

```
                     /note="Correia element; hmmfs hit to HMM Correia (1 - 62),
                     score: 78.71"
                     /label=Correia
     repeat_unit     562. .604
                     /note="Correia element; hmmfs hit to HMM Correia (114 -
                     156), score: 51.92"
                     /label=Correia
     misc_feature    complement(667. .676)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(668. .1813)
                     /gene="dapE"
     CDS             complement(668. .1813)
                     /gene="dapE"
                     /EC_number="3.5.1.18"
                     /note="NMA1730, dapE, succinyl-diaminopimelate
                     desuccinylase, len: 381aa; similar to many eg. SW:P24176
                     (DAPE_ECOLI) dapE, succinyl-diaminopimelate desuccinylase
                     from Escherichia coli (375 aa) fasta scores; E(): 0, 56.9%
                     identity in 371 aa overlap. Contains Pfam match to entry
                     PF01546 Peptidase_M20."
                     /codon_start=1
                     /transl_table=11
                     /product="succinyl-diaminopimelate desuccinylase"
                     /protein_id="CAB84958.1"
                     /db_xref="GI:7380372"
                     /db_xref="SPTREMBL:Q9JTL0"
                     /translation="MTETQSLELAKALISRPSVTPDDRDCQKLLVERLYKIGFAAEEL
                     HFGDTKNIWLRRGTKVPVVCFAGHTDVVPTGPVEKWDSPPFEPTERDGRLYGRGAADM
                     KTSIACFVTACERFVAEHPDHQGSIALLITSDEEGDALDGTTKVVDVLKARGELIDYC
                     IVGEPTAVDKLGDMIKNGRRGSLSGNLTVKGKQGHIAYPHLAINPVHTFAPALLELTQ
                     EVWDEGNKYFPPTSFQISNINGGTGATNVIPGELNVKFNFRFSTESTEAGLKQRVHAI
                     LDKHGVQYDLQWSCSGQPFLTQAGKLTDVARAAIAETCGIEAELSTTGGTSDGRFIKA
                     IAKELIELGPSNATIHQINENVRLDDIPKLSAVYEGILARLLAGNAV"
     misc_feature    complement(836. .1786)
                     /gene="dapE"
                     /note="Pfam match to entry PF01546 Peptidase_M20, , score
                     252.50, E-value 5.7e-72"
     misc_feature    complement(1863. .1872)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(1927. .2547)
                     /gene="NMA1731"
     CDS             complement(1927. .2547)
                     /gene="NMA1731"
                     /note="NMA1731, conserved hypothetical protein, len:
                     206aa; similar to SW:P44013 (Y552_HAEIN) hypothetical
                     protein from Haemophilus influenzae (207 aa) fasta scores;
                     E(): 0, 59.6% identity in 203 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAB84959.1"
                     /db_xref="GI:7380373"
                     /db_xref="SPTREMBL:Q9JQZ3"
                     /translation="MLTPKSCDLFNIPFFQFSQLKKYQPESIPQIKADYKENWQIWQQ
                     LIQQVAAELGAPFAPPHIERWCNGWQVRAHFFAYFKYEQYKNSAAILSILLNRRRLSV
                     SLDWHCYKADVSPIALPDYNRWLDNFDTEKYASFDMWHGAESEYDDYRTVAQQNESDR
                     KLQNDEDFFCIGKHIERDDLGRQDVAKWIAETVEDLLPLYEACHGK"
     RBS             complement(2551. .2558)
     misc_feature    complement(2570. .2579)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(2581. .3090)
                     /gene="NMA1732"
     CDS             complement(2581. .3090)
                     /gene="NMA1732"
                     /note="NMA1732, conserved hypothetical protein, len:
                     169aa; similar to SW:P43971 (Y243_HAEIN) hypothetical
                     protein from Haemophilus influenzae (172 aa) fasta scores;
                     E(): 0, 49.7% identity in 169 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAB84960.1"
                     /db_xref="GI:7380374"
                     /db_xref="SPTREMBL:Q9JTK9"
                     /translation="MNPFETQSVTFAEPIEMLYACHGKVRRFCGQVAMLSDYIAENGC
                     NQLVLQTIRQITRYFNVAAPLHHEDEEENFFPLLLQYAPQAQESVDELLRQHVSLHGN
                     WDAVAAEFAKLEADNAYIPDAEAFKRFVAGYDVHLAIEEPLFDMGKMFIPKEKLTEIG
                     EIMAARRRR"
     gene            complement(3185. .3736)
                     /gene="NMA1733"
     CDS             complement(3185. .3736)
                     /gene="NMA1733"
                     /note="NMA1733, H.8 outer membrane protein, len: 183aa;
                     strongly similar to SW:P07212 (H8_NEIME) H.8 outer
                     membrane protein from Neisseria meningitidis (183 aa)
                     fasta scores; E(): 0, 96.7% identity in 183 aa overlap.
                     Contains Pfam match to entry PF00127 copper-bind, Copper
                     binding proteins, plastocyanin/azurin family and Prosite
                     match to PS00196 Type-1 copper (blue) proteins signature."
                     /codon_start=1
                     /transl_table=11
                     /product="H.8 outer membrane protein"
                     /protein_id="CAB84961.1"
                     /db_xref="GI:7380375"
                     /db_xref="SWISS-PROT:P57025"
                     /translation="MKAYLALISAAVIGLAACSQEPAAPAAEATPAAEAPASEAPAAE
                     AAPADAAEAPAAGNCAATVESNDNMQFNTKDIQVSKACKEFTITLKHTGTQPKTSMGH
                     NIVIGKTEDMDGIFKDGVGAADTDYVKPDDARVVAHTKLIGGGEESSLTLDPAKLADG
                     EYKFACTFPGHGALMNGKVTLVD"
     misc_feature    complement(3191. .3568)
                     /gene="NMA1733"
                     /note="Pfam match to entry PF00127 copper-bind, Copper
                     binding proteins, plastocyanin/azurin family, score
                     218.60, E-value 9.1e-62"
     misc_feature    complement(3212. .3259)
                     /gene="NMA1733"
                     /note="PS00196 Type-1 copper (blue) proteins signature."
     gene            complement(4017. .4439)
                     /gene="NMA1734"
     CDS             complement(4017. .4439)
                     /gene="NMA1734"
                     /note="NMA1734, unknown, len: 140aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB84962.1"
                     /db_xref="GI:7380376"
                     /db_xref="SPTREMBL:Q9JTK8"
                     /translation="MNLEQLGRRDALLSGLLKQAGQWRRLDAAVKKLLPANLHPHFQT
                     ACIEDGRLVLLAANNMAASRLKMIAPSVLPQLAGLDASIRSVSVRLVPKPEKPPKTNT
                     LHLSKAALESFGSAAVKLEKRHPELAEALANLVRRHGA"
     misc_feature    4306. .4315
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(4486. .4495)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            4570. .7320
                     /gene="secA"
     CDS             4570. .7320
                     /gene="secA"
                     /note="NMA1735, secA, preprotein translocase SecA subunit,
                     len: 916aa; similar to many eg. SW:P10408 (SECA_ECOLI)
                     preprotein translocase SecA subunit from Escherichia coli
                     (901 aa) fasta scores; E(): 0, 58.0% identity in 920 aa
                     overlap. Contains Pfam match to entry PF01043
                     SecA_protein, SecA protein, amino terminal region and
                     Prosite match to PS01312 Protein secA signatures."
                     /codon_start=1
                     /transl_table=11
                     /product="preprotein translocase SecA subunit"
                     /protein_id="CAB84963.1"
                     /db_xref="GI:7380377"
```

```
                     /db_xref="SPTREMBL:Q9JTK7"
                     /translation="MLTNIAKKIFGSRNDRLLKQYRKSVARINALEEQMQALSDADLQ
                     AKTAEFKQRLADGQTLDGILPEAFAVCREASRRTLGMRHFDVQLIGGMVLHDGKIAEM
                     RTGEGKTLVATLAVYLNALAGKGVHVVTVNDYLASRDAGIMEPLYNFLGLTVGVIISD
                     MQPFDRQNAYAADITYGTNNEFGFDYLRDNMVTDQYDKVQRELNFAVVDEVDSILIDE
                     ARTPLIISGQADDNIQLYQIMNTVPPHLVRQETEEGEGDYWVDEKAHQVILSEAGHEH
                     AEQILTQMGLLAENDSLYSAANIALMHHLMAALRAHTLFHKDQHYVIQDGEIVIVDEF
                     TGRLMSGRRWSEGLHQAVEAKEGVEIKRENQTLASITFQNYFRLYTKLSGMTGTADTE
                     AFEFQSIYNLETVIIPTNRPVQRKDFNDQIFRSAEEKFEAVVKDIEECHKRGQPVLVG
                     TTSIENSELVSHLLQKAGLPHNVLNAKEHEREALIVAQAGKVGAITVATNMAGRGTDI

  Query Match              2.7%;  Score 45.4;  DB 1;  Length 326301;
  Best Local Similarity   55.4%;  Pred. No. 3.1;
  Matches  108;  Conservative    0;  Mismatches   86;  Indels    1;  Gaps    1;

Qy     538 agcaccggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaagga 597
           ||||| |||   || | |||| ||||   |||  | |  || ||   ||  || || |||
Db 158403 AGCACTGGAATTCGCCCTGTCTACCGCCGCCAAGAAAAAGGCAGACCGCGAACACATGGA 158462

Qy     598 -gtaggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaa 656
            ||  ||| |  | |||| |||| |||| |  |   ||      |  | | | |   |
Db 158463 CGTGCGCGTCCAAATCAACCGCGACACCGGCGAATACCAAACCTTCCGCCGCTGGCTGAT 158522

Qy     657 tgccgccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcagg 716
           || |||||| |    | |   |  | | ||| |||  |||| ||||   |       ||
Db 158523 TGTCGCCGATGAAGACTATACCTATCCCGATGTCGAAAAAACCATCGAGGAAATCCAAGA 158582

Qy     717 agaaattgtcggcgc 731
           ||||||  |||| |
Db 158583 GGAAATTCCCGGCAC 158597


RESULT  41
AX044033
LOCUS       AX044033               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 112 from Patent WO0066791.
ACCESSION   AX044033
VERSION     AX044033.1  GI:11342917
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 112 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    87189 a  93501 c  84627 g  84663 t
ORIGIN


  Query Match              2.7%;  Score 45.4;  DB 6;  Length 349980;
  Best Local Similarity   55.4%;  Pred. No. 3.1;
  Matches  108;  Conservative    0;  Mismatches   86;  Indels    1;  Gaps    1;

Qy     538 agcaccggacaacggcttgtcgaccgtttccacaataccggtagtatgctgacgcaagga 597
           ||||| |||   || | |||| ||||   |||  | |  || ||   ||  || || |||
Db 205925 AGCACTGGAATTCGCCCTGTCTACCGCCGCCAAGAAAAAGGCAGACCGCGAACACATGGA 205984

Qy     598 -gtaggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaa 656
            ||  ||| |  | |||| |||| |||| |  |   ||      |  | | | |   |
Db 205985 CGTGCGCGTCCAAATCAACCGCGACACCGGCGAATACCAAACCTTCCGCCGCTGGCTGAT 206044

Qy     657 tgccgccgaagctttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcagg 716
           || |||||| |    | |   |  | | ||| |||  |||| ||||   |       ||
Db 206045 TGTCGCCGATGAAGACTATACCTATCCCGATGTCGAAAAAACCATCGAGGAAATCCAAGA 206104

Qy     717 agaaattgtcggcgc 731
           ||||||  |||| |
Db 206105 GGAAATTCCCGGCAC 206119


RESULT  42
SCVALSFP/c
LOCUS       SCVALSFP               3619 bp    DNA     linear   BCT 03-MAR-1998
DEFINITION  S.coelicolor valS, fpgs, ndk genes.
ACCESSION   Y13070
VERSION     Y13070.1  GI:2125894
KEYWORDS    folylpolyglutamate synthase; fpgs gene; NDK gene; nucleoside
            diphosphate kinase; ORFX; valS gene; valyl-tRNA synthetase.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 3619)
  AUTHORS   Burger,A., Brandt,B., Susstrunk,U., Thompson,C.J. and Wohlleben,W.
  TITLE     Analysis of a Streptomyces coelicolor A3(2) locus containing the
            nucleoside diphosphate kinase (ndk) and folylpolyglutamate
            synthetase (folC) genes
  JOURNAL   FEMS Microbiol. Lett. 159 (2), 283-291 (1998)
  MEDLINE   98164371
REFERENCE   2  (bases 1 to 3619)
  AUTHORS   Burger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1997) A. Burger, Univeristaet Tuebingen,
            Lehrstuhl Mikrobiologie/Biotechnologie, Auf der Morgenstelle 28,
            D-72076, FRG
FEATURES             Location/Qualifiers
     source          1. .3619
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="M145"
                     /db_xref="taxon:100226"
     gene            1. .779
                     /gene="valS"
     CDS             <1. .779
                     /gene="valS"
                     /codon_start=3
                     /transl_table=11
                     /product="valyl-tRNA synthetase"
                     /protein_id="CAA73510.1"
                     /db_xref="GI:2125895"
                     /db_xref="SWISS-PROT:O06851"
                     /translation="ILSRLNSVVAEVDAYYEDYQFANVSDALFHFAWDEVFDWYVELS
                     KTTFQAGGEAAEVSKRVLGEVLDVTLRLLHPVVPFVTETLWTTLTGGESVVIAEWPTD
                     SGFRDAGAEREIETVQSVITEVRRFRADQGLQPGQRVPARLTLAGSALAAHEAAVRQL
                     LRLQPEGDAFTATATLPVAGVEVALDLSGVIDFAAERKRLAKDLAAAEKEKAQANAKL
                     GNEAFLAKAPDQVVDKIRGRLAKADEDITRITAQLEKLPE"
     gene            1254. .2588
                     /gene="fpgs"
     CDS             1254. .2588
                     /gene="fpgs"
                     /EC_number="6.3.2.17"
                     /codon_start=1
                     /transl_table=11
                     /product="folylpolyglutamate synthase"
                     /protein_id="CAA73511.1"
                     /db_xref="GI:2125896"
                     /db_xref="SPTREMBL:O08416"
                     /translation="MDAALRAVETELAGRWGETKLEPSVTRIAALMDVLGEPQRSYPS
                     IHITGTNGKTSTARMIEALLGAFDLRTGRYTSPHVQSITERISLDGAPISAERFIETY
                     EDIKPYVEMVDAQRSTGCSFFEVLTGMAYAAFADAPVDVAVVEVGMGGSWDATNVIDG
                     DVAVVTPIDLDHTDRLGGTPGEIAVEKAGIVKQDATVILAQQPVDAAQVMLKKAVDVN
```

```
                     ATVAREGLEFGVVAREVPSAGSWPPCAGLGGEYTEVYLPLHGAHQAHNAAVALAAVEA
                     FFGVGAQRPEPLDVDTVRKAFASVASPGRMEVVRRSPTVVLDAAHNPAGARVTAEAVG
                     EAFQFSRLIGVVGASADKNVRGLLEAFEPVFAEVVVTQNSSHRAMDADELAAVAVEVF
                     GDDRVQVEPRLPDALEAAITLAEEEGEFAGGGVLVTGSVITVGEARLLLRRG"
     CDS             2595. .2948
                     /note="orfX"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAA73512.1"
                     /db_xref="GI:2125897"
                     /db_xref="SPTREMBL:O06852"
                     /translation="MRTLCASTLIGEFFVIFAGLVAMKDPDLSMGTVWTVSGIAMVLS
                     VLLCGVITRPGGIALGWALQIALIAAGFVVPTMFFLGAIFAALWWASVHYGRKIDEAK
                     ARFAAQAEASAPDPA"
     gene            3073. .3486
                     /gene="ndk"
     CDS             3073. .3486
                     /gene="ndk"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleoside diphosphate kinase"
                     /protein_id="CAA73513.1"
                     /db_xref="GI:2125898"
                     /db_xref="SWISS-PROT:P50589"
                     /translation="MTQRSLVLLKPEAVRRGLTGEIISRIERKAGWQITALELRTLDT
                     ETLEQHYGEHKGKPFYEPLVEFMASGPVVAMVVEGERVIEGVRALAGPTDPIAAAPGS
                     IRGDYGVIVRENLIHASDSEESAERELKIFFPGRV"
BASE COUNT      524 a   1329 c   1275 g    491 t
ORIGIN


  Query Match             2.5%;  Score 42.2;  DB 1;  Length 3619;
  Best Local Similarity   49.5%;  Pred. No. 8;
  Matches  138;  Conservative    0;  Mismatches  138;  Indels    3;  Gaps    1;

Qy     292 gggcacgaagtccattccccccttcgacaaccatgcctcacattccgattctgatgaagcc 351
           || || || | |  |  |  | | |     || |||||   | |  |||    |||   |||
Db    1853 GGCCAGGATGACGGTGGCGTCCTGCTTGACGATGCCGGCCTTCTCGACCGCGATCTCGCC 1794

Qy     352 ggtagtcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccat 411
           ||  ||||||  ||     || |  |   ||   || || |||  |  |||  |||
Db    1793 GGGCGTCCCGCCGAGCCGGTCGGTGTGGTCCAGGTCGATGGGGGTGACGACGGCCACGTC 1734

Qy     412 cccgccgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagg 471
           |||| ||| | |  ||  ||     |  || | || | ||||||  | | | | ||   |
Db    1733 CCCGTCGATGACGTTGGTGGCGTCCCAGGAGCCGCCCATCCCGACCTCGACGACG---GC 1677

Qy     472 gatatatacagctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgac 531
            |  |  || | | || |    ||||||  |  |   ||  | | |   ||||    ||
Db    1676 CACGTCCACGGGTGCGTCGGCGAAGGCGGCGTACGCCATGCCGGTCAGCACCTCGAAGAA 1617

Qy     532 aaccgcagcaccggacaacggcttgtcgaccgtttccac 570
               |      |  | ||  ||||||| | || ||
Db    1616 GGAACAGCCGGTACTCCTCTGCGCGTCGACCATCTCGAC 1578


RESULT  43
SCC88
LOCUS       SCC88                   42816 bp    DNA     linear   BCT 15-FEB-2000
DEFINITION  Streptomyces coelicolor cosmid C88.
ACCESSION   AL139298
VERSION     AL139298.1  GI:6983731
KEYWORDS    50S ribosomal protein L21; 50S ribosomal protein L27;
            dehydrogenase; fpgS; glycosyltransferase; integral membrane
            protein; integrase; membrane protein; mreB; mreC; mreD; ndk; obg;
            pbp2; secreted protein; sfr; transferase; valS.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 42816)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 42816)
  AUTHORS   Saunders,D. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 42816)
  AUTHORS   Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
            program of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid C88 Lies between and overlaps with cosmids C123 and C80 on
            the AseI-C genomic restriction fragment.
FEATURES                Location/Qualifiers
     source             1. .42816
                        /organism="Streptomyces coelicolor A3(2)"
                        /strain="A3(2)"
                        /db_xref="taxon:100226"
                        /clone="cosmid C88"
     gene               1. .3409
                        /gene="SCC88.01"
     misc_feature       1. .106
                        /gene="SCC88.01"
                        /note="nominal overlap with S. coelicolor cosmid StC123"
     CDS                1. .3409
                        /partial
                        /gene="SCC88.01"
                        /note="Protein sequence is in conflict with the conceptual
                        translation; SCC88.01 possible glycosyltransferase, len:
                        >1135 aa; N-terminal region similar to TR:Q9XC92
                        (EMBL:AF146532) Klebsiella pneumoniae glycosyltransferase,
                        329 aa; fasta scores: opt: 214 z-score: 245.4 E(): 3e-06;
                        21.3% identity in 263 aa overlap and C-terminal region
                        similar to SW:TAGF_BACSU (EMBL:X15200) Bacillus subtilis
                        teichoic acid biosynthesis protein F TagF, 746 aa; fasta
                        scores: opt: 1056 z-score: 1203.0 E(): 0; 41.5% identity
                        in 381 aa overlap"
```

```
                     /codon_start=2
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="CAB75371.1"
                     /db_xref="GI:6983732"
                     /translation="MRKENAGLGAARNTGMTRLAPESEYVAFVDSDDMIPPDAYRLMV
                     ESLDESGSDFVTGNVQHINSTKIWQSPMHRILAGGAVRRTHISRNKKLLTDRIACNKV
                     FRRSFWEKHDLSFPEGVLYEDVPVILPAQFLAEAVDIISEPTYYWRLREGEAAPSITQ
                     RRTEPKAVRDRAAAVESVSRFFASRPGESGAELKHAYDHTALTGDLRIFLNVLPDGDA
                     EFRAEFLRVINKYLDQVDPRVVMELPATARIKWLLVRKHALDDLVDLIKAERRGDAVK
                     ISGLVRKYVSYPTVEAAAAGLPKKARRIDPDLRLHAPLQDIRWDGGRLRLSGHAWVDQ
                     IDQSGRRSAAKVLVLKKDGSRRATVVPLRNTHVPEATSLSGRKHNNYDWSGWAADLDP
                     LKLRRGSQWEEGTWRVGIGIATAGLLRKRAVQVKGPTRANHPPYQWLDDDFRLLPTPV
                     NGALRLTVERVRALITGHRAVGADIEIDGEIRVPLGAGESVSLRVTNKKSGERLHYPV
                     TLSAADGAHTPFTVRVPMRDVALLPEHQDVVKEENAAGGADPAAASRRNWSTQLVAVA
                     PDGGERRFSTVVREGLADGEMRLPASLAENADRNEIALVSGHNGYLKFSGRPLQAKIT
                     SVRWTGDRYVVEGVSAADIDGADFVVKARDRFDEKTVALRTSPDGSFAAEFTPASMSG
                     PDGTLPLKSGRWNFFLRTPDGHDIPFLVDRLAVDGFPVLGAAKDREYEFESRWYNFPQ
                     LHCPSDLSVLERGAYRQERLRKDVYEAGRTQPLRDAVLFISYNGKQYSDSPRAMHEEL
                     LRRGADLEYLWLVRDGQVDLPDTAQKVRFWSTEWYEALARCRYIVTNAHLPHWMERRP
                     GQVIVQTWHGTMLKKIGLDIEAPKFDPQYHERLIQEARNWTMLVSSNRFSTPILKRAM
                     GFDGKIIETGYPRNDYLYAPDLAERAREIKERLGVPSDKRVILYAPTWRDDLSHRRGQ
                     FKFDLRIDVEDARARLGDDHVLLIRRHSNVVDSIPGAGNGFVHDVSEYPDIADLYLAA
                     DVLVTDYSSVMFDYAHLKRPMLFFTYDLEHYRDKLRGFYFDFENDAPGPLIRTSEELV
                     SALGDLDKVSAEYADRYRRFQELFCDLDDGRAAARVVDLMLEQAREV"
stem_loop            3416. .3460
gene                 3580. .5310
                     /gene="SCC88.02"
CDS                  3580. .5310
                     /gene="SCC88.02"
                     /note="SCC88.02. possible secreted protein, len: 576 aa;
                     similar to TR:CAB51964 (EMBL:AL109661) Streptomyces
                     coelicolor hypothetical 100.7 kD protein SC6E10.10, 943
                     aa; fasta scores: opt: 522 z-score: 556.2 E(): 1.4e-23;
                     31.8% identity in 471 aa overlap. Contains possible
                     N-terminal region signal peptide sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secreted protein"
                     /protein_id="CAB75372.1"
                     /db_xref="GI:6983733"
                     /translation="MFTSRARLRTRRSRLLTCTALAVAAGMLVTTPAVAADKPSPGIE
                     VPTPKSQLPKLDRAQPERKQAPRMRAAATPAPRFDVDGDGFSDMLLRLQDGSTQVGLT
                     TADGFTPYTINVDDYEERQKDVITPGDLDGNGGPEILTLSSTGTLSLFQSWGADSTGY
                     ATWSGTGWQIYNKVFAPGDVTGDGRPDLLARTYGGDLYLYQGTGSVDSPFRGRVKVGG
                     GWEMFDQLVGMGDMNGDGIGDVVARTRLGYLYFYAGKGSATAPLATRDLLGEGWTGYN
                     QIIAADDWDGDGLGDLLGRTRSGDLYYYRADGTGNFEAKQQAGTGFQPVEFFAGSGVY
                     PHFGKASLVGMDTRGTLYWYGVKNDGQLTGREQISDTGGWSGARVAYASSLDKDGYAD
                     LLELYDGVLYNYAATADESLATGWGKYNLFLGPGDLNDDGKGDLLARDTSGTLYLFRG
                     KGHGASLSSPLKVGGGWNAYDKLVGAGDFTGDGRTDIVARTPAGKLYLYRGTGVSTSP
                     FAGKVDLGSGWQQYNKLASPGDIDGDGRADLLAVNSGGDLYRYTSYTDGRFKARVKLG
                     TGWNTYRNLY"
stem_loop            5327. .5362
gene                 complement(5387. .8215)
                     /gene="SCC88.03c"
CDS                  complement(5387. .8215)
                     /gene="SCC88.03c"
                     /note="SCC88.03c, possible transferase, len: 942 aa;
                     N-terminal region similar to hypothetical proteins, e. g.
                     TR:O69853 (EMBL:AL023702) Streptomyces coelicolor
                     hypothetical 66.0 kD protein SC1C3.11, 586 aa; fasta
                     scores: opt: 1214 z-score: 1395.6 E(): 0; 40.7% identity
                     in 573 aa overlap and C-terminal region similar to
                     TR:O69854 (EMBL:AL023702) Streptomyces coelicolor putative
                     transferase SC1C3.12, 697 aa; fasta scores: opt: 658
                     z-score: 754.0 E(): 0; 30.1% identity in 574 aa overlap.
                     Contains Pfam match at N-terminal region to entry PF00534
                     Glycos_transf_1, Glycosyl transferases group 1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transferase"
                     /protein_id="CAB75373.1"
                     /db_xref="GI:6983734"
                     /translation="MKITFLLTWGDEMGGTEQAVYTQATHLAPRHDIEVLSVFKTREQ
                     PFFSVDERVSVRYLVDRTGKTERPVRESDLTAEDCRHLAGLPSDLISPKWEATFDALS
                     DVEMQRALRTIDTDVLITTSPALMSAVADLAPSRVITIQQEHRPSQLRGGTGEPLLLR
                     APAIDALVVLTERTKQWLEESLGKAAPRLAAIPNAIPEGFRPRSSLTGKTIVMPRRLV
                     PDKQVDHAIQAFAKALPDHPGWRLRIFGDGPQMSRLRNLIQGLGLHDSVELLGPSQHM
                     TEEWARASLTILPSQDGEAFPLVLLEAFAAGVPAVAYDIVTGPAEIIRHGEDGLLVPP
                     NDVESLAEAISRLMGDEALLRSYGEKAHEGSTRFAADVIVKQWEELFTELVSRRDDPR
                     RMAERADRIAHRVAHGGAGRFHAAVAADRTAPSSGDQRAREVVIGASDRSLVRAGGRL
                     SEVRDDLQGSEIVQRNFETVVEALESSGIPYVLLRDRDDNPRRRLAVDAAEQTRVRKA
                     LAGAYEGKAVYAELLKPRTHAPGVLLAERLEAVGEVAGLRVFRPVVTSTRTLRFGPAY
                     GCDIEFWRQVPEEEGGDGQFVAPLRPSAVGPKLPSLTPDARTRVKDREYPTLEPLTRK
                     LVSDITFPVDAVYTWVDDSDPRWQERRARRRAALGLEAESSGDEAARFRNRDELRYSL
                     RSLAMFAPWIRKIYLVTDDQTPEWLNTEHEGIEVVSHRDIFTDQDCLPTFNSHSIESQ
                     LHHIDGLSEQFLYLNDDVFIGRPVGAQRFFLPNGASRFFWSPTTVPVGEPTEEDEGYF
                     AAAKNNRALLEERFGVTVANSFVHAPHPLRRSVLESIEEDFPESVARTAATPMRGWQD
                     ISMVSSLHHHYGYLTGASVPSSIRCAYIDVGTYSRHPELTRLLAMRGHDVFCLGESQD
                     AEVPEHEQARIVEAFLRAYFPVKSPYER"
misc_feature         complement(7124. .7570)
                     /gene="SCC88.03c"
                     /note="Pfam match to entry PF00534 Glycos_transf_1,
                     Glycosyl transferases group 1, score 99.70, E-value
                     2.1e-27"
gene                 complement(8222. .9211)
                     /gene="SCC88.04c"
CDS                  complement(8222. .9211)
                     /gene="SCC88.04c"
                     /note="SCC88.04c, unknown, len: 329 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SCC88.04c"
                     /protein_id="CAB75374.1"
                     /db_xref="GI:6983735"
                     /translation="MKQLPTPHQVMRKGYAAVRSARSRLDSLVLVPPSPDDWPRVAAD
                     HLSLLDGRTLNVCVTVPDDLTAASASLVLGGGSSPVVIPLTLVQRGDGRVEARGTSVV
                     DLLENVRADSPPVPAGMRRFLLEAGQWRLSVVFTDVRGQERRFALAAAPRLMPDGPTL
                     PEPVRADGTYCRVIASSTGRAYLSLGRDHAEAEIVSVDIGWSQITLRGRLVNSPAEAC
                     SGDVELIRRNGKVSRTLPATWQGDVFTCTVHTADFGASGSKEQIWDVRLRRPRGRSLK
                     LCRRRTDVRSPGDVFRMPNRLLTADDGTALRINPYYTPVGSLAFRAALIPSGS"
gene                 complement(9211. .11019)
                     /gene="SCC88.05c"
```

```
  Query Match              2.5%;  Score 42.2;  DB 1;  Length 42816;
  Best Local Similarity   49.5%;  Pred. No. 14;
  Matches  138;  Conservative    0;  Mismatches  138;  Indels    3;  Gaps    1;

Qy     292 gggcacgaagtccattccccccttcgacaaccatgcctcacattccgattctgatgaagcc 351
           || || || | |  |  | |  | |     || ||||  | |  |||   |||   |||
Db   38303 GGCCAGGATGACGGTGGCGTCCTGCTTGACGATGCCGGCCTTCTCGACCGCGATCTCGCC 38362

Qy     352 ggtagtcccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccat 411
           ||  ||||||  ||     || |  |   ||  || || |||  |   |||  |||
Db   38363 GGGCGTCCCGCCGAGCCGGTCGGTGTGGTCCAGGTCGATGGGGGTGACGACGGCCACGTC 38422

Qy     412 cccgccgacggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagg 471
           |||| ||| | |  ||  ||     |  || | || | ||||||  | | | | ||   |
Db   38423 CCCGTCGATGACGTTGGTGGCGTCCCAGGAGCCGCCCATCCCGACCTCGACGACG---GC 38479

Qy     472 gatatatacagctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgac 531
            |  |  || | | || |    ||||||  |  |   ||  | | |   ||||    ||
Db   38480 CACGTCCACGGGTGCGTCGGCGAAGGCGGCGTACGCCATGCCGGTCAGCACCTCGAAGAA 38539

Qy     532 aaccgcagcaccggacaacggcttgtcgaccgtttccac 570
             |           |  | ||  |||||||  | || ||
Db   38540 GGACAGCCGGTACTCCTGCTGCGCGTCGACCATCTCGAC 38578


RESULT  44
AC102254/c
LOCUS       AC102254               70423 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP24-181A10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC102254
VERSION     AC102254.1  GI:17061340
KEYWORDS    HTG; HTGS_PHASE0.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 70423)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-181A10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70423)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L18267
                Center clone name: 181_A_10
            --------------.
            * NOTE: This record contains 90 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      674: contig of 674 bp in length
            *      675 774: gap of      100 bp
            *      775     1472: contig of 698 bp in length
            *     1473 1572: gap of      100 bp
            *     1573     2265: contig of 693 bp in length
            *     2266 2365: gap of      100 bp
            *     2366     3060: contig of 695 bp in length
            *     3061 3160: gap of      100 bp
            *     3161     3834: contig of 674 bp in length
            *     3835 3934: gap of      100 bp
            *     3935     4612: contig of 678 bp in length
            *     4613 4712: gap of      100 bp
            *     4713     5394: contig of 682 bp in length
            *     5395 5494: gap of      100 bp
            *     5495     6171: contig of 677 bp in length
            *     6172 6271: gap of      100 bp
            *     6272     6954: contig of 683 bp in length
            *     6955 7054: gap of      100 bp
            *     7055     7749: contig of 695 bp in length
            *     7750 7849: gap of      100 bp
            *     7850     8538: contig of 689 bp in length
            *     8539 8638: gap of      100 bp
            *     8639     9325: contig of 687 bp in length
            *     9326 9425: gap of      100 bp
            *     9426    10132: contig of 707 bp in length
            *    10133 10232: gap of      100 bp
            *    10233    10931: contig of 699 bp in length
            *    10932 11031: gap of      100 bp
            *    11032    11689: contig of 658 bp in length
            *    11690 11789: gap of      100 bp
            *    11790    12463: contig of 674 bp in length
            *    12464 12563: gap of      100 bp
            *    12564    13228: contig of 665 bp in length
            *    13229 13328: gap of      100 bp
            *    13329    14013: contig of 685 bp in length
            *    14014 14113: gap of      100 bp
            *    14114    14800: contig of 687 bp in length
            *    14801 14900: gap of      100 bp
            *    14901    15600: contig of 700 bp in length
            *    15601 15700: gap of      100 bp
            *    15701    16397: contig of 697 bp in length
            *    16398 16497: gap of      100 bp
            *    16498    17157: contig of 660 bp in length
            *    17158 17257: gap of      100 bp
            *    17258    17944: contig of 687 bp in length
            *    17945 18044: gap of      100 bp
            *    18045    18717: contig of 673 bp in length
            *    18718 18817: gap of      100 bp
            *    18818    19518: contig of 701 bp in length
            *    19519 19618: gap of      100 bp
            *    19619    20285: contig of 667 bp in length
            *    20286 20385: gap of      100 bp
            *    20386    21060: contig of 675 bp in length
            *    21061 21160: gap of      100 bp
            *    21161    21835: contig of 675 bp in length
            *    21836 21935: gap of      100 bp
            *    21936    22604: contig of 669 bp in length
            *    22605 22704: gap of      100 bp
            *    22705    23367: contig of 663 bp in length
            *    23368 23467: gap of      100 bp
            *    23468    24145: contig of 678 bp in length
            *    24146 24245: gap of      100 bp
            *    24246    24939: contig of 694 bp in length
            *    24940 25039: gap of      100 bp
            *    25040    25722: contig of 683 bp in length
            *    25723 25822: gap of      100 bp
            *    25823    26488: contig of 666 bp in length
            *    26489 26588: gap of      100 bp
            *    26589    27282: contig of 694 bp in length
            *    27283 27382: gap of      100 bp
            *    27383    28075: contig of 693 bp in length
            *    28076 28175: gap of      100 bp
            *    28176    28851: contig of 676 bp in length
            *    28852 28951: gap of      100 bp
            *    28952    29608: contig of 657 bp in length
            *    29609 29708: gap of      100 bp
            *    29709    30386: contig of 678 bp in length
            *    30387 30486: gap of      100 bp
            *    30487    31159: contig of 673 bp in length
            *    31160 31259: gap of      100 bp
            *    31260    31949: contig of 690 bp in length
            *    31950 32049: gap of      100 bp
            *    32050    32752: contig of 703 bp in length
            *    32753 32852: gap of      100 bp
            *    32853    33549: contig of 697 bp in length
            *    33550 33649: gap of      100 bp
            *    33650    34260: contig of 611 bp in length
            *    34261 34360: gap of      100 bp
            *    34361    34999: contig of 639 bp in length
```

```
*    35000 35099: gap of      100 bp
*    35100    35801: contig of 702 bp in length
*    35802 35901: gap of      100 bp
*    35902    36595: contig of 694 bp in length
*    36596 36695: gap of      100 bp
*    36696    37401: contig of 706 bp in length
*    37402 37501: gap of      100 bp
*    37502    38182: contig of 681 bp in length
*    38183 38282: gap of      100 bp
*    38283    38951: contig of 669 bp in length
*    38952 39051: gap of      100 bp
*    39052    39717: contig of 666 bp in length
*    39718 39817: gap of      100 bp
*    39818    40501: contig of 684 bp in length
*    40502 40601: gap of      100 bp
*    40602    41287: contig of 686 bp in length
*    41288 41387: gap of      100 bp
*    41388    42086: contig of 699 bp in length
*    42087 42186: gap of      100 bp
*    42187    42884: contig of 698 bp in length
*    42885 42984: gap of      100 bp
*    42985    43632: contig of 648 bp in length
*    43633 43732: gap of      100 bp
*    43733    44422: contig of 690 bp in length
*    44423 44522: gap of      100 bp
*    44523    45221: contig of 699 bp in length
*    45222 45321: gap of      100 bp
*    45322    46018: contig of 697 bp in length
*    46019 46118: gap of      100 bp
*    46119    46780: contig of 662 bp in length
*    46781 46880: gap of      100 bp
*    46881    47539: contig of 659 bp in length
*    47540 47639: gap of      100 bp
*    47640    48318: contig of 679 bp in length
*    48319 48418: gap of      100 bp
*    48419    49095: contig of 677 bp in length
*    49096 49195: gap of      100 bp
*    49196    49883: contig of 688 bp in length
*    49884 49983: gap of      100 bp
*    49984    50679: contig of 696 bp in length
*    50680 50779: gap of      100 bp
*    50780    51488: contig of 709 bp in length
*    51489 51588: gap of      100 bp
*    51589    52278: contig of 690 bp in length
*    52279 52378: gap of      100 bp
*    52379    53051: contig of 673 bp in length
*    53052 53151: gap of      100 bp
*    53152    53848: contig of 697 bp in length
*    53849 53948: gap of      100 bp

  Query Match             2.5%;  Score 42.2;  DB 2;  Length 70423;
  Best Local Similarity   36.3%;  Pred. No. 15;
  Matches  136;  Conservative    0;  Mismatches  238;  Indels    1;  Gaps    1;

Qy   1217 caaacaaacgccacccgaagaccaaagtgccgtttgacggtaaagggtttccgaattttg 1276
          || |   | ||||  |   || | |  ||   |  |  | ||   | |||  |  ||
Db  65100 CAGAAGTAAGCCAAACTTTGAGCTAGCTGTTCTAGGGTGCTATCTGATTTATGTGTTAAT 65041

Qy   1277 aaaaagacgtaaaatacgatacgagaattaataccgctgtaccacaagtgaatcctatag 1336
          | |   | | |||| | |  | ||  |   |||    |  ||||| |||| ·     ||
Db  65040 AGATGCAAGGAAAACAAGGAAGGAAGAGAGATAGAATTCCACCACCAGTGGCGATGATTC 64981

Qy   1337 atgaacccgtctttaatcctaaaggttctgtcggatcggctcattcttggtctataactg 1396
          |
Db  64980 AGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 64921

Qy   1397 ccagaattcaatacgcaaaattaccaaggcaaggtagaatcagatatatcccacctaaaa 1456
                                                    | |||||  ||  || |
Db  64920 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGTATATTACATGTATAT 64861

Qy   1457 attactctccttcagcaccgctaccaaaaggacct-aataatggatatttggataaattt 1515
           ||  | |   |  | | |  ||  ||||||   | ||||   ||||   | |||  |
Db  64860 GTTTATATGTGTATGTATCAATAATAAAAGGTTATGAATATGAGATAGGGGAATATGTGA 64801

Qy   1516 ggtaatgaatggactaaaggtccatcaagaactaaaggtcaagaatttgaatgggatgtt 1575
          |||| |  | |||  ||| |    |    || | | |   |   | |  | ||   | |
Db  64800 GGTATTTCAGGGAGGAAATGAAGGTATGAAAATTATGCCAATACAGTGTACTGATGTATG 64741

Qy   1576 caattgtctaaaaca 1590
           |||| || |||| |
Db  64740 AAATTATCAAAAATA 64726


RESULT  45
AP004215
LOCUS       AP004215               98980 bp    DNA     linear   HTG 27-SEP-2001
DEFINITION  Oryza sativa chromosome 7 clone OJ1429_G07, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP004215
VERSION     AP004215.1  GI:15787872
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1429_G07.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 98980)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1429_G07
  JOURNAL   Published Only in Database (2001) In press
REFERENCE   2  (bases 1 to 98980)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .98980
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="7"
                     /clone="OJ1429_G07"
BASE COUNT    28747 a  20334 c  20154 g  29695 t     50 others
ORIGIN


  Query Match             2.5%;  Score 42.2;  DB 2;  Length 98980;
  Best Local Similarity   49.3%;  Pred. No. 16;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy   1449 acctaaaaattactctccttcagcaccgctaccaaaaggacctaataatggatatttgga 1508
          || ||| |   | | | |    |   |   |||||    |   |   | ||  || ||
Db  86070 ACTTAACATGGAATTTACAATGGAGGCATAACCAACCAAAAAAACATAAGGTGATCTGCT 86129

Qy   1509 taaatttggtaatgaatggactaaaggtccatcaagaactaaaggtcaagaatttgaatg 1568
          |||| |||  ||| | | || |||    |  |  |   |||||   ||   |  | |||
Db  86130 TAAAGTTGAAAATAACTAGATTAAGTAACACTATAAGTCTAAAATACATATAGCTTAATT 86189

Qy   1569 ggatgttcaattgtctaaaacaggaagagagcaacttggatgggctagtagggatggtaa 1628
          |   | |        || ||  |  ||      |||| | | ||||||    |||||| |
```

```
Db  86190 GTTCGATTTGGCAATTAGAATGGTCAGTCTAAGACTTTGTTTGGCTAGATTAGATGGTTA 86249

Qy   1629 gcatttaaatatatcaattgatggaaagattacacacaaatga 1671
          ||       ||| |   ||   |||||| ||   || | || ||
Db  86250 GCCAGATTATACACTTATATTTGGAAAAATAGAACTCTAAAGA 86292


Search completed: June 30, 2002, 09:25:29
Job time: 58147 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 07:52:06 ; Search time 27321.5 Seconds
(without alignments)
1445.323 Million cell updates/sec

Title: US-09-303-518D-463
Perfect score: 1887
Sequence: 1 ttgggcatttcccgcaaaat.....catatcccatttatgaatag 1887

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1886	99.9	1887	6	A96430	A96430 Sequence 46
2	1427.8	75.7	1827	6	AX044009	AX044009 Sequence
3	1427.8	75.7	11796	1	AE002419	AE002419 Neisseria
4	1427.8	75.7	349980	6	AX044030	AX044030 Sequence
5	1231.2	65.2	1671	6	A96432	A96432 Sequence 46
6	1231.2	65.2	311321	1	NMA3Z2491	AL162754 Neisseria
7	1198.8	63.5	2304	6	AX236465	AX236465 Sequence
8	1198.8	63.5	2304	6	AX239751	AX239751 Sequence
9	1196	63.4	2028	6	AX236421	AX236421 Sequence
10	1196	63.4	2028	6	AX239731	AX239731 Sequence
11	1196	63.4	4425	6	AX236407	AX236407 Sequence
12	1196	63.4	4425	6	AX239715	AX239715 Sequence
13	1194	63.3	2019	6	AX236447	AX236447 Sequence
14	1194	63.3	2019	6	AX239733	AX239733 Sequence
15	1194	63.3	2256	6	AX236451	AX236451 Sequence
16	1194	63.3	2256	6	AX236459	AX236459 Sequence
17	1194	63.3	2256	6	AX239737	AX239737 Sequence
18	1194	63.3	2256	6	AX239745	AX239745 Sequence
19	1194	63.3	2421	6	AX236449	AX236449 Sequence
20	1194	63.3	2421	6	AX236453	AX236453 Sequence
21	1194	63.3	2421	6	AX239735	AX239735 Sequence
22	1194	63.3	2421	6	AX239739	AX239739 Sequence
23	642.8	34.1	683	6	A96426	A96426 Sequence 45
24	322.4	17.1	381	6	A96424	A96424 Sequence 45
c 25	283.4	15.0	11700	1	AE002393	AE002393 Neisseria
c 26	283.4	15.0	349980	6	AX044029	AX044029 Sequence
27	269.6	14.3	195767	1	NMA7Z2491	AL162758 Neisseria
28	226.6	12.0	10057	1	AE002420	AE002420 Neisseria
29	119.8	6.3	1497	6	A96138	A96138 Sequence 17
30	119.8	6.3	1744	6	A61829	A61829 Sequence 9
31	119.8	6.3	3287	6	A61821	A61821 Sequence 1
32	119.8	6.3	3294	1	AF142582	AF142582 Neisseria
c 33	108.2	5.7	340806	1	NMA1Z2491	AL162752 Neisseria
34	105.2	5.6	1449	6	A96134	A96134 Sequence 16
35	105	5.6	1452	6	A96132	A96132 Sequence 16
36	105	5.6	10869	1	AE002559	AE002559 Neisseria
37	105	5.6	172325	6	AX044035	AX044035 Sequence
38	96	5.1	375	6	A96130	A96130 Sequence 16
c 39	80.2	4.3	7218	6	I66494	I66494 Sequence 14
40	49	2.6	125020	9	AF429315	AF429315 Homo sapi
c 41	48.4	2.6	125020	9	AF429315	AF429315 Homo sapi
c 42	45.8	2.4	41755	2	AC006179	AC006179 Homo sapi
43	45.2	2.4	390	6	AR135147	AR135147 Sequence
44	45.2	2.4	115890	2	AP002824	AP002824 Homo sapi
c 45	45.2	2.4	157078	9	AC096727	AC096727 Homo sapi
46	44.6	2.4	30941	1	SC4C6	AL079355 Streptomy
47	44.2	2.3	10598	1	AE006028	AE006028 Caulobact
c 48	44.2	2.3	12026	1	AE005109	AE005109 Halobacte
49	44.2	2.3	169022	2	AC027354	AC027354 Homo sapi
50	43.8	2.3	605	8	AF377524	AF377524 Zea mays
51	43.8	2.3	605	8	AF377525	AF377525 Zea mays
52	43.8	2.3	605	8	AF377527	AF377527 Zea mays
53	43.8	2.3	605	8	AF377536	AF377536 Zea mays
54	43.8	2.3	605	8	AF377542	AF377542 Zea mays
55	43.6	2.3	291	3	AF134687	AF134687 Plasmodiu
56	43.6	2.3	204652	2	PFMAL13P6	AL049183 Plasmodiu
57	43.4	2.3	231	3	AF134664	AF134664 Plasmodiu
58	43.4	2.3	131535	9	AC005875	AC005875 citb_188_
59	43.4	2.3	133954	9	AL355493	AL355493 Human DNA
60	43.4	2.3	204881	2	AL591111	AL591111 Homo sapi
61	43.2	2.3	674	3	PFA310620	AJ310620 Plasmodiu
62	43.2	2.3	674	3	PFA310621	AJ310621 Plasmodiu
63	43.2	2.3	674	3	PFA310623	AJ310623 Plasmodiu
c 64	43.2	2.3	347660	1	AP002994	AP002994 Mesorhizo
65	43	2.3	942	1	AF169562	AF169562 Mesorhizo
66	42.6	2.3	3151	1	AB055506	AB055506 Streptomy
67	42.4	2.2	15311	1	AE007164	AE007164 Mycobacte
68	42.2	2.2	605	8	AF377520	AF377520 Zea mays
69	42.2	2.2	605	8	AF377521	AF377521 Zea mays
70	42.2	2.2	605	8	AF377522	AF377522 Zea mays
71	42.2	2.2	605	8	AF377523	AF377523 Zea mays
72	42.2	2.2	605	8	AF377526	AF377526 Zea mays
73	42.2	2.2	605	8	AF377529	AF377529 Zea mays

	74	42.2	2.2	605	8	AF377530	AF377530 Zea mays
	75	42.2	2.2	605	8	AF377531	AF377531 Zea mays
	76	42.2	2.2	605	8	AF377532	AF377532 Zea mays
	77	42.2	2.2	605	8	AF377533	AF377533 Zea mays
	78	42.2	2.2	605	8	AF377537	AF377537 Zea mays
	79	42.2	2.2	605	8	AF377538	AF377538 Zea mays
	80	42.2	2.2	605	8	AF377540	AF377540 Zea mays
	81	42.2	2.2	1696	8	ZMU14599	U14599 Zea mays Bz
	82	42	2.2	47852	1	MTV023	AL022022 Mycobacte
	83	42	2.2	315000	1	RME603644	AL603644 Rhizobium
	84	41.8	2.2	329	3	AF134691	AF134691 Plasmodiu
	85	41.8	2.2	816	6	E03078	E03078 gDNA encodi
	86	41.8	2.2	1450	1	FVBNAM	D90316 Flavobacter
c	87	41.8	2.2	20701	1	AE006958	AE006958 Mycobacte
	88	41.6	2.2	3977	1	GCGLUABCD	X81191 C.glutamicu
c	89	41.6	2.2	106650	9	AC007708	AC007708 Homo sapi
c	90	41.4	2.2	10911	1	AE007300	AE007300 Sinorhizo
	91	41.4	2.2	185556	9	AC097519	AC097519 Homo sapi
	92	41.2	2.2	9960	1	AE006085	AE006085 Pasteurel
c	93	41	2.2	39446	1	SCE87	AL132674 Streptomy
	94	40.8	2.2	2653	1	XXGDH	X16399 Gene for gl
c	95	40.8	2.2	47573	3	AF030694	AF030694 Plasmodiu
	96	40.6	2.2	10947	1	AE002514	AE002514 Neisseria
	97	40.6	2.2	326301	1	NMA6Z2491	AL162757 Neisseria
	98	40.6	2.2	349980	6	AX044033	AX044033 Sequence
	99	40.4	2.1	942	1	AF169564	AF169564 Mesorhizo
	100	40.4	2.1	36224	1	SCD78	AL034355 Streptomy

ALIGNMENTS

```
RESULT   1
A96430
LOCUS       A96430                  1887 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 463 from Patent WO9924578.
ACCESSION   A96430
VERSION     A96430.1  GI:6780107
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1887)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 463 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1887
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      554 a    468 c    493 g    371 t      1 others
ORIGIN
```

Query Match 99.9%; Score 1886; DB 6; Length 1887;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1887; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 60

Qy      61 catgcacacgcctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcag 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAG 120

Qy     121 catttcgaacccgacgggaaataccacctattcggcagcaggggggagcttgccnagcgc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGC 180

Qy     181 aacggccatatcggattgggaaacatacaaagccatcagttgggccacctgatgattcaa 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AACGGCCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAA 240

Qy     241 caggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaa 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAA 300

Qy     301 ttccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtccc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCC 360

Qy     361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 420

Qy     421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 480

Qy     481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 540

Qy     541 accggacaacggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagta 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTA 600

Qy     601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 660

Qy     661 gccgaagccttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAA 720

Qy     721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatg 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATG 780

Qy     781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 840

Qy     841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900

Qy     901 aatgccgcacaaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaa 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAA 960

Qy     961 gggattggagctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAG 1020

Qy    1021 cggtcgcagatgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 1080

Qy    1081 gccgatgcggcatacgccaaatacccgtccccttaccattcccgaaatatccgttcaaac 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 1140

Qy    1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggc 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 1200

Qy    1201 aaaaatgtcaaactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAA 1260

Qy    1261 gggtttccgaattttgagaagcacgtgaaatatgatacgaagctcgatattcaagaatta 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1261 GGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATTA 1320

Qy   1321 tcgggggggcggtatacctaaggctaagcctgtgtttgatgcgaaaccgagatgggaggtt 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTT 1380

Qy   1381 gataggaagcttaataaattgacaactcgtgagcaggtggagaaaaatgttcaggaaacg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACG 1440

Qy   1441 agaagaaggagtcagagtagtcagtttaaagcccatgcgcaacgagaatgggaaaataaa 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAAATAAA 1500

Qy   1501 acagggttagattttaatcattttataggtggtgatatcaataagaaaggcacagtaaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ACAGGGTTAGATTTTAATCATTTTATAGGTGGTGATATCAATAAGAAAGGCACAGTAACA 1560

Qy   1561 ggagggcatagtctaacccgtggtgatgtacgggtgatacaacaaacctcggcacctgat 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GGAGGGCATAGTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTGAT 1620

Qy   1621 aaacatggggtttatcaagcgacagtggaaattaaaaagcctgatggaagttgggaggtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AAACATGGGGTTTATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAGTTGGGAGGTG 1680

Qy   1681 aaaacgaaaaaaggtgggaaagtgatgaccaagcacaccatgttcccaaaagattgggat 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AAAACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCCAAAAGATTGGGAT 1740

Qy   1741 gaggctagaattagggctgaagttacttcggcttgggaaagtagaataatgcttaaggat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GAGGCTAGAATTAGGGCTGAAGTTACTTCGGCTTGGGAAAGTAGAATAATGCTTAAGGAT 1800

Qy   1801 ataaaatggcagggtacaagtaaatcgggtattaaaatagaaggatttaccgaacctaat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AATAAATGGCAGGGTACAAGTAAATCGGGTATTAAAATAGAAGGATTTACCGAACCTAAT 1860

Qy   1861 agaacagcatatcccatttatgaatag 1887
          |||||||||||||||||||||||||||
Db   1861 AGAACAGCATATCCCATTTATGAATAG 1887


RESULT   2
AX044009
LOCUS       AX044009               1827 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 88 from Patent WO0066791.
ACCESSION   AX044009
VERSION     AX044009.1  GI:11342894
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1827)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 88 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .1827
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
BASE COUNT      539 a    446 c    454 g    387 t      1 others
ORIGIN


  Query Match           75.7%;  Score 1427.8;  DB 6;  Length 1827;
  Best Local Similarity  95.6%;  Pred. No. 0;
  Matches 1468;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 60

Qy     61 catgcacacgcctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcag 120
          |||||||||||||||||||||||||| | ||||||||||||||||||||||||||||||
Db     61 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 120

Qy    121 catttcgaacccgacgggaaataccacctattcggcagcaggggggagcttgccnagcgc 180
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||| |||||
Db    121 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGC 180

Qy    181 aacggccatatcggattgggaaacatacaaagccatcagttgggccacctgatgattcaa 240
          | |||||||||||||||||||| ||||||||||||||||||| ||||||||||||||||
Db    181 AGCGGCCATATCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAA 240

Qy    241 caggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaa 300
          ||||||||| || ||||||||||||||||||||||||||||||||||||||||||||| ||
Db    241 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 300

Qy    301 ttccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtccc 360
          ||||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    301 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 360

Qy    361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTTGACGGATTTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 420

Qy    421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 480

Qy    481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 540

Qy    541 accggacaacggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagta 600
          |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db    541 ACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTA 600

Qy    601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 660

Qy    661 gccgaagccttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    661 GCCGAAGCCTTCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAA 720

Qy    721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatg 780
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    721 ATTGTCGGCGCAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATG 780

Qy    781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 840

Qy    841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAA 960

Qy    961 gggattggagctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
          |||||||||||||| ||||||||||||||||||||||||||||||||||| |||||||||
Db    961 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAG 1020

Qy   1021 cggtcgcagatgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 1080

Qy   1081 gccgatgcggcatacgccaaatacccgtccccttaccattcccgaaatatccgttcaaac 1140
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 1140

Qy    1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggc 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 1200

Qy    1201 aaaaatgtcaaactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAA 1260

Qy    1261 gggtttccgaattttgagaagcacgtgaaatatgatacgaagctcgatattcaagaatta 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATTA 1320

Qy    1321 tcgggggcggtatacctaaggctaagcctgtgtttgatgcgaaaccgagatgggaggtt 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTT 1380

Qy    1381 gataggaagcttaataaattgacaactcgtgagcaggtggagaaaaatgttcaggaaacg 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAATA 1440

Qy    1441 agaagaaggagtcagagtagtcagtttaaagcccatgcgcaacgagaatgggaaaataaa 1500
           || |   | | |   |  ||| | ||||     ||||| |||| |||  |||||| |||
Db    1441 AGGAACGGTAATATAAACAGTAACTTTAGCCAACATGCTCAACTAGAGAGGGAAATTAAT 1500

Qy    1501 acagggttagattttaatcattttataggtggtgat 1536
           | |     |  |    || |  |||    ||  |||
Db    1501 AAACTAAAATCTGCCGATGAAATTAATTTTGCAGAT 1536


RESULT   3
AE002419
LOCUS       AE002419               11796 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 61 of 206 of
            the complete genome.
ACCESSION   AE002419 AE002098
VERSION     AE002419.1  GI:7225863
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 11796)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 11796)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11796
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            complement(174. .1139)
                     /gene="NMB0635"
     CDS             complement(174. .1139)
                     /gene="NMB0635"
                     /note="similar to SP:P37247 PID:1197006 percent identity:
                     61.94; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase, IS30 family"
                     /protein_id="AAF41058.1"
                     /db_xref="GI:7225864"
                     /translation="MSYTQLTQGERYHIQYLSRHCTVTEIAKQLNRHKSTISREIRRH
                     RTQGQQYSAEKAQRQSRTIKQRKRQPYKLDSQLIQHIDPLIRRKLSPEQVCAYLCKHH
                     QITLHHSTIYRYLRQDKSNGSTLWQHLRICSKPYRKRYGSTWTRGKVPNRVGIENRPA
                     IVDQKSRIGDWEADTIVGKGQKSALLTLVERVTRYTIICKLDSLKAEDTARAAVRALK
                     AHKDRVHTITMDNGKEFYQHTKITKALKAETYFCRPYHSWEKGLNENTNGLIRQYFPK
                     QTDFRNISDREIRRVQDELNHRPRKTLGYETPSVLFLNLFQPLIH"
     gene            complement(1303. .1686)
                     /gene="NMB0636"
     CDS             complement(1303. .1686)
                     /gene="NMB0636"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41059.1"
                     /db_xref="GI:7225865"
                     /translation="MTDTDTQADRFEQMMWQAVDKLFEQHDGKLESMDGREQELVLIW
                     RTEADIGNGSILQFVCNWGFPAAEKTCSVLKKIGAVHSAMLIHRAADALDKEIRRLQS
                     EGKNLKEMWDITSRQQNRLTAEQSG"
     gene            complement(1744. .3120)
                     /gene="NMB0637"
     CDS             complement(1744. .3120)
                     /gene="NMB0637"
                     /note="similar to PID:1652362 percent identity: 67.97;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="argininosuccinate lyase"
                     /protein_id="AAF41060.1"
                     /db_xref="GI:7225866"
                     /translation="MHDKTWSGRFNEPVSELVKQYTASIGFDRRLAEWDIQGSLAHAQ
                     MLKETGVLDEGDLADIRRGMAEILEEIRSGKIEWSSDLEDVHMNIERRLTDKIGDAGK
                     RLHTGRSRNDQVATDIRLWLRDQITVIQSLIQSLQTALLDLAEQNAETVMPGFTHLQV
                     AQPVSFGHHMLAYVEMLGRDNERMADCRCRVNRMPLGAAALAGTTYPIQREITAELLG
                     FEQICQNSLDAVSDRDFAIEFTAAASLVMVHLSRLSEELILWMSPRFGFIDIADRFCT
                     GSSIMPQKKNPDVPELVRGKSGRVIGHLIGLITLMKSQPLAYNKDNQEDKEPLFDTAD
                     TLIDTLRIYADMMRGVTVKPDNMRAAVMQGFATATDLADYLVKKGMPFRDAHEVVAQA
                     VRHADQAGVDLSELPLEVLQGFSDLIADDVYGVLTPEGSLNARNHLGGTAPEQVRFQV
                     KRWREMLA"
     gene            complement(3139. .4008)
                     /gene="NMB0638"
     CDS             complement(3139. .4008)
                     /gene="NMB0638"
                     /note="similar to PID:1628575 percent identity: 74.13;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP--glucose-1-phosphate uridylyltransferase"
                     /protein_id="AAF41061.1"
                     /db_xref="GI:7225867"
                     /translation="MKPIRKAVFPVAGMGTRFLPATKASPKEMLPIVDKPLIQYAVEE
                     AVEAGCTEMVFVTGRNKRSIEDHFDKAYELETELEMRHKDKLLEHVRNILPPNITCLY
                     IRQAEALGLGHAVLCARAAIGDEPFAVILADDLIDAPKGALKQMVEVYGRSGNSILGV
                     ETVEASQTGSYGIVETEQLKQFQRITGIVEKPKPEDAPSNLAVVGRYILTPRIFDLLT
                     NLPRGAGNEIQLTDGIAKLLDHEFVLAHPFEGTRYDCGSKLGYLEATVAYGLKHPETG
                     EPFRRLLEKYRTE"
```

```
gene            complement(4036. .4635)
                /gene="NMB0639"
CDS             complement(4036. .4635)
                /gene="NMB0639"
                /note="conserved hypothetical protein; identified by
                Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF41062.1"
                /db_xref="GI:7225868"
                /translation="MSEKPEKIVLASGNAGKLEEFGNLFKPYSITVLPQSAFGIPECP
                EPYPTFVENALAKARHAAKYSGLPALADDSGICAAALNGAPGIHSARYAGDNPKSDTA
                NNLKLAAELVGKADKSCCYVCVLVFVRHKDDPRPIIAEGVWHGQWNDTPLGQNGFGYD
                PYFYLPEHGKTAAELDTEVKNRESHRAQALAELLRKLAL"
gene            complement(4628. .4861)
                /gene="NMB0640"
CDS             complement(4628. .4861)
                /gene="NMB0640"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF41063.1"
                /db_xref="GI:7225869"
                /translation="MNRRKIYLLSVALFTLAFMLLVLLGAYLLTVGSKAFAVASFLFA
                FGALFGQIGSLALYLRHKSLRAAQSATKENRYV"
gene            complement(4969. .5502)
                /gene="NMB0641"
CDS             complement(4969. .5502)
                /gene="NMB0641"
                /note="similar to GB:L42023 SP:P44529 PID:1003152
                PID:1222038 PID:1204383 percent identity: 92.61;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="inorganic pyrophosphatase"
                /protein_id="AAF41064.1"
                /db_xref="GI:7225870"
                /translation="MADFNQILTPGDVDGGIINVVNEIPAGSNHKIEWNRKLAAFQLD
                RVEPAIFAKPTNYGFIPQTLDEDGDELDVLLVTEQPLATGVFLEARVIGVMKFVDDGE
                VDDKIVCVPADDRNNGNAYKTLSDLPQQLIKQIEFHFNMYKDLKKAGTTKVESWGDAE
                EAKKVIKESIERWNKQA"
gene            complement(5603. .6061)
                /gene="NMB0642"
CDS             complement(5603. .6061)
                /gene="NMB0642"
                /note="similar to GB:D10165 SP:P24236 PID:581152
                PID:669113 PID:912430 percent identity: 58.45; identified
                by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="dATP pyrophosphohydrolase"
                /protein_id="AAF41065.1"
                /db_xref="GI:7225871"
                /translation="MPMAKPLKYPVSALVVLYSGDGGILLIERTHPEGFWQSVTGSLE
                PGETVAQTARREVWEETGILLADGQLQDWHDSTVYEIYHHWRHRYPKGVFENREHLFS
                AEIPRDTPIALQPEEHVSYGWFDMEEAAEKVFSPSNRRAILELGRFLGKR"
tRNA            6272. .6348
                /gene="tRNA-Pro-3"
                /product="tRNA-Pro"
gene            6272. .6348
                /gene="tRNA-Pro-3"
gene            6506. .8332
                /gene="NMB0643"
CDS             6506. .8332
                /gene="NMB0643"
                /note="MafB-related protein; identified by Oxford;
                putative"
                /codon_start=1
                /transl_table=11
                /product="MafB-related protein"
                /protein_id="AAF62313.1"
                /db_xref="GI:7413435"
                /translation="MGISRKISLILSILAVCLPMHAHASDLANDSFIRQVLDRQHFEP
                DGKYHLFGSRGELAERSGHIGLGKIQSHQLGNLMIQQAAIKGNIGYIVRFSDHGHEVH
                SPFDNHASHSDSDEAGSPVDGFSLYRIHWDGYEHHPADGYDGPQGGGYPAPKGARDIY
                SYDIKGVAQNIRLNLTDNRSTGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDRSG
                NAAEAFNGTADIVKNIIGAAGEIVGAGDAVQGISEGSNIAVMHGLGLLSTENKMARIN
                DLADMAQLKDYAAAAIRDWAVQNPNAAQGIEAVSNIFMAAIPIKGIGAVRGKYGLGGI
                TAHPIKRSQMGAIALPKGKSAVSDNFADAAYAKYPSPYHSRNIRSNLEQRYGKENITS
                STVPPSNGKNVKLADQRHPKTGVPFDGKGFPNFEKHVKYDTKLDIQELSGGGIPKAKP
                VFDAKPRWEVDRKLNKLTTREQVEKNVQEIRNGNINSNFSQHAQLEREINKLKSADEI
                NFADGMGKFTDSMNDKAFSRLVKSVKENGFTNPVVEYVEINGKAYIVRGNNRVFAAEY
                LGRTHELKFKKVDFPVPNTSWKNPTDVLNESGNVKRPRYRSK"
gene            8335. .8799
                /gene="NMB0644"
```

Query Match 75.7%; Score 1427.8; DB 1; Length 11796;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 1468; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy      1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6506 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 6565

Qy     61 catgcacacgcctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcag 120
          |||||||||||||||||||||||||| | ||||||||||||||||||||||||||||||
Db   6566 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 6625

Qy    121 catttcgaacccgacgggaaataccacctattcggcagcagggggagcttgccnagcgc 180
          |||||||||||||||||||||||||||||||||||||||||||||| |||||| |||||
Db   6626 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGC 6685

Qy    181 aacggccatatcggattgggaaacatacaaagccatcagttgggccacctgatgattcaa 240
          | |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||
Db   6686 AGCGGCCATATCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAA 6745

Qy    241 caggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaa 300
          ||||||||| || ||||||||||||||||||||||||||||||||||||||||||| ||
Db   6746 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 6805

Qy    301 ttccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtccc 360
           |||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   6806 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 6865

Qy    361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   6866 GTTGACGGATTTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 6925

Qy    421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6926 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 6985

Qy    481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6986 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 7045

Qy    541 accggacaacggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagta 600
          |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db   7046 ACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTA 7105

Qy    601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7106 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 7165

Qy    661 gccgaagccttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   7166 GCCGAAGCCTTCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAA 7225

Qy    721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatg 780
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   7226 ATTGTCGGCGCAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATG 7285

Qy    781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
```

```
Db   7286 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 7345

Qy    841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
Db   7346 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 7405

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaa 960
Db   7406 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAA 7465

Qy    961 gggattggagctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
Db   7466 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAG 7525

Qy   1021 cggtcgcagatgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
Db   7526 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 7585

Qy   1081 gccgatgcggcatacgccaaataccgtccccttaccattcccgaaatatccgttcaaac 1140
Db   7586 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 7645

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggc 1200
Db   7646 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 7705

Qy   1201 aaaaatgtcaaactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaa 1260
Db   7706 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAA 7765

Qy   1261 gggtttccgaattttgagaagcacgtgaaatatgatacgaagctcgatattcaagaatta 1320
Db   7766 GGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATTA 7825

Qy   1321 tcgggggggcggtatacctaaggctaagcctgtgtttgatgcgaaaccgagatgggaggtt 1380
Db   7826 TCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTT 7885

Qy   1381 gataggaagcttaataaattgacaactcgtgagcaggtggagaaaaatgttcaggaaacg 1440
Db   7886 GATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAATA 7945

Qy   1441 agaagaaggagtcagagtagtcagtttaaagcccatgcgcaacgagaatgggaaaataaa 1500
Db   7946 AGGAACGGTAATATAAACAGTAACTTTAGCCAACATGCTCAACTAGAGAGGGAAATTAAT 8005

Qy   1501 acagggttagattttaatcattttataggtggtgat 1536
Db   8006 AAACTAAAATCTGCCGATGAAATTAATTTTGCAGAT 8041


RESULT   4
AX044030
LOCUS       AX044030               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 109 from Patent WO0066791.
ACCESSION   AX044030
VERSION     AX044030.1  GI:11342914
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 109 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.-seq 1: 1 to
                     349980 349980 bases-seq 108: 300001 to 649980 349980
                     bases-seq 109: 600001 to 949980 349980 bases-seq 110:
                     900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
                     349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
                     113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    82523 a  82940 c  96712 g  87805 t
ORIGIN


  Query Match             75.7%;  Score 1427.8;  DB 6;  Length 349980;
  Best Local Similarity   95.6%;  Pred. No. 0;
  Matches 1468;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy       1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
Db   75614 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 75673

Qy      61 catgcacacgcctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcag 120
Db   75674 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 75733

Qy     121 catttcgaacccgacgggaaataccacctattcggcagcagggggggagcttgccnagcgc 180
Db   75734 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGC 75793

Qy     181 aacggccatatcggattgggaaacatacaaagccatcagttgggccacctgatgattcaa 240
Db   75794 AGCGGCCATATCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAA 75853

Qy     241 caggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaa 300
Db   75854 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 75913

Qy     301 ttccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtccc 360
Db   75914 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 75973

Qy     361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
Db   75974 GTTGACGGATTTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 76033

Qy     421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
Db   76034 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 76093

Qy     481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
Db   76094 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 76153

Qy     541 accggacaacggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagta 600
Db   76154 ACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTA 76213

Qy     601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
Db   76214 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 76273

Qy     661 gccgaagccttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
Db   76274 GCCGAAGCCTTCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAA 76333

Qy     721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatg 780
Db   76334 ATTGTCGGCGCAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATG 76393

Qy     781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
Db   76394 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 76453

Qy     841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
```

```
Db  76454 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 76513

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaa 960
Db  76514 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAA 76573

Qy    961 gggattggagctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
Db  76574 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAG 76633

Qy   1021 cggtcgcagatgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
Db  76634 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 76693

Qy   1081 gccgatgcggcatacgccaaatacccgtcccccttaccattcccgaaatatccgttcaaac 1140
Db  76694 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 76753

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggc 1200
Db  76754 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 76813

Qy   1201 aaaaatgtcaaactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaa 1260
Db  76814 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAA 76873

Qy   1261 gggtttccgaattttgagaagcacgtgaaatatgatacgaagctcgatattcaagaatta 1320
Db  76874 GGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATTA 76933

Qy   1321 tcgggggggcggtatacctaaggctaagcctgtgtttgatgcgaaaccgagatgggaggtt 1380
Db  76934 TCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTT 76993

Qy   1381 gataggaaagcttaataaattgacaactcgtgagcaggtggagaaaaatgttcaggaaacg 1440
Db  76994 GATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAATA 77053

Qy   1441 agaagaaggagtcagagtagtcagtttaaagcccatgcgcaacgagaatgggaaaataaa 1500
Db  77054 AGGAACGGTAATATAAACAGTAACTTTAGCCAACATGCTCAACTAGAGAGGGAAATTAAT 77113

Qy   1501 acagggttagattttaatcattttataggtggtgat 1536
Db  77114 AAACTAAAATCTGCCGATGAAATTAATTTTGCAGAT 77149


RESULT   5
A96432
LOCUS       A96432                 1671 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 465 from Patent WO9924578.
ACCESSION   A96432
VERSION     A96432.1  GI:6780108
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1671)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 465 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1671
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      484 a    440 c    402 g    345 t
ORIGIN


  Query Match             65.2%;  Score 1231.2;  DB 6;  Length 1671;
  Best Local Similarity   96.3%;  Pred. No. 2.7e-281;
  Matches 1260;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy      1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
Db      1 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 60

Qy     61 catgcacacgcctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcag 120
Db     61 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 120

Qy    121 catttcgaacccgacgggaaataccacctattcggcagcagggggagcttgccnagcgc 180
Db    121 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGC 180

Qy    181 aacggccatatcggattgggaaacatacaaagccatcagttgggccacctgatgattcaa 240
Db    181 AGCGGTCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCAACCTGTTCATCCAG 240

Qy    241 caggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaa 300
Db    241 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 300

Qy    301 ttccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtccc 360
Db    301 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 360

Qy    361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
Db    361 GTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 420

Qy    421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
Db    421 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 480

Qy    481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
Db    481 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 540

Qy    541 accggacaacggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagta 600
Db    541 ACCGGACAACGGCTTGTCGACCGTTTCCACAATACCGGTAGTATGCTGACGCAAGGAGTA 600

Qy    601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
Db    601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 660

Qy    661 gccgaagccttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
Db    661 GCCGAAGCTTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAA 720

Qy    721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatg 780
Db    721 ATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTTATG 780

Qy    781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
Db    781 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 840

Qy    841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
Db    841 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900

Qy    901 aatgccgcacaaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaa 960
Db    901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTACGGCAGTCATCCCCGTCAAA 960

Qy    961 gggattggagctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
Db    961 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAG 1020

Qy   1021 cggtcgcagatgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
```

```
Db   1021 CGGTCGCAGATGGGCGAGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 1080

Qy   1081 gccgatgcggcatacgccaaataccgtccccttaccattcccgaaatatccgttcaaac 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 1140

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGA 1200

Qy   1201 aaaaatgtcaaactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaa 1260
          || ||||| |||||||||| || ||||||||||||||||    || ||||||||||||||
Db   1201 AAGAATGTGAAACTGGCAAACAAACGCCACCCGAAGACCAAAGTGCCGTTTGACGGTAAA 1260

Qy   1261 gggtttccgaattttgagaagcacgtgaaatatgatacgaagctcgata 1309
          ||||||||||||||| ||  |||| ||||| |||||||   |  |||
Db   1261 GGGTTTCCGAATTTTGAAAAAGACGTAAAATACGATACGAGAATTAATA 1309


RESULT   6
NMA3Z2491
LOCUS       NMA3Z2491             311321 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 3/7.
ACCESSION   AL162754 AL157959
VERSION     AL162754.2  GI:7379424
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 311321)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 311321)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .311321
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     RBS             115. .120
     gene            129. .587
                     /gene="NMA0698"
     CDS             129. .587
                     /gene="NMA0698"
                     /note="NMA0698, possible periplasmic protein, len: 152 aa;
                     unknown, contains a probable N-terminal signal sequence
                     and lies within a region of unusually low GC content"
                     /codon_start=1
                     /transl_table=11
                     /product="putative periplasmic protein"
                     /protein_id="CAB83985.1"
                     /db_xref="GI:7379425"
                     /db_xref="SPTREMBL:Q9JQQ2"
                     /translation="MLTKLKILLFLFLFVFVLAINLLFFFFSSDIESFGNYQFEYVYD
                     KGWPANYILVMKDGNEGNFDKIISGLVLEYYKEDDNIYFSYIDGQGFASDSCYYKPEI
                     LYGKIILNKNHIININSMEKNNFLSEDKIMKGTRNWLADPKNKCNIQTLD"
     gene            742. .1077
                     /gene="NMA0699"
     CDS             742. .1077
                     /gene="NMA0699"
                     /note="NMA0699, pseudogene, probable ABC transporter
                     protein, len: 336 bp; similar to C-termini of many ABC
                     transporters e.g. TR:Q46973 (EMBL:U47048), mtfB,
                     Escherichia coli microcin transport protein (707 aa),
                     fasta scores; E(): 2.6e-19, 53.2% identity in 111 aa
                     overlap. Contains PS00211 ABC transporters family
                     signature. NMA0686 may be the remainder of this
                     pseudogene"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="putative ABC transporter protein (pseudogene)"
     misc_feature    868. .912
                     /gene="NMA0699"
                     /note="PS00211 ABC transporters family signature"
     gene            complement(1101. .2327)
                     /gene="NMA0700"
     CDS             complement(1101. .2327)
                     /gene="NMA0700"
                     /note="NMA0700, possible ribonuclease BN, len: 408 aa;
                     similar to SW:RBN_ECOLI (EMBL:L19201), rbn, Escherichia
                     coli ribonuclease BN (EC 3.1.-.-), fasta scores; E():
                     2.3e-28, 34.3% identity in 268 aa overlap. Longer than rbn
                     at the C-terminus. Also similar to TR:O85449
                     (EMBL:AF067083) Vitreoscilla sp. hypothetical protein (376
                     aa), fasta scores; E(): 0, 45.0% identity in 333 aa
                     overlap. Contains a region similar to NMA0677, fasta
                     scores; E(): 4.3e-08, 64.4% identity in 45 aa overlap.
                     Contains hydrophobic, possible membrane-spanning regions,
                     as does rbn"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribonuclease BN"
                     /protein_id="CAB83987.1"
                     /db_xref="GI:7379426"
                     /db_xref="SPTREMBL:Q9JVT9"
                     /translation="MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLL
                     ALVPVLTVMVAVASIFPVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTA
                     IGSVMLVVTSLMLIRTIDNTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFM
                     VGSVQDAALASGAPQWSGALRTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATA
                     FCLETARSLFTWYMGNFDGYRSIYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQ
                     GEAFRRGFDSRGRFDDVLKILLLLDAAQKEGKALPVQEFRRHINMGYDELGELLEKLA
                     RHGYIYSGRQGWVLKTGADSIELNELFKLFVYRPLPVERDHVNQAVDAVMMPCLQTLN
                     MTLAEFDAQAKKQQQS"
     misc_feature    1376. .1385
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(2417. .2426)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2444. .2453
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            2486. .2743
                     /gene="NMA0701"
     CDS             2486. .2743
                     /gene="NMA0701"
                     /note="NMA0701, possible pseudogene, len: 258 bp; shows
                     weak similarity to part of SW:WRBA_ECOLI (EMBL:M99166),
                     wrbA, Escherichia coli Trp repressor binding protein (197
                     aa), fasta scores; E(): 0.064, 29.3% identity in 82 aa
                     overlap. Also similar to part of TR:O85450 (EMBL:AF067083)
                     Vitreoscilla sp. Trp repressor binding protein (fragment)
                     (124 aa), fasta scores; E(): 6.6e-14, 56.6% identity in 83
                     aa overlap"
                     /codon_start=1
                     /pseudo
```

```
                /transl_table=11
                /product="hypothetical protein NMA0701 (pseudogene)"
misc_feature    complement(2755. .2764)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    2778. .2787
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
RBS             2861. .2864
gene            2871. .3530
                /gene="NMA0702"
CDS             2871. .3530
                /gene="NMA0702"
                /note="NMA0702, len: 219 aa; similar to TR:Q52925
                (EMBL:Z50189), exsB, Rhizobium meliloti putative regulator
                of succinoglycan biosynthesis (not a transcriptional
                regulator) (234 aa), fasta scores; E(): 3.1e-15, 32.7%
                identity in 205 aa overlap. Also similar to many
                hbacterial ypothetical proteins e.g. SW:YBAX_HAEIN
                (EMBL:U32798), HI1191, Haemophilus influenzae hypothetical
                protein (196 aa), fasta scores; E(): 0, 79.2% identity in
                173 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0702"
                /protein_id="CAB83989.1"
                /db_xref="GI:7379427"
                /db_xref="SPTREMBL:Q9JVT8"
                /translation="MSNQKALVIFSGGQDSTTCLIQAIQTYGRENVQAITFQYGQRHA
                VELERARWIAQDLGVKQTVLDLSLMRQITHNALMDDTAAIETAENGVPNTFVDGRNAL
                FLLYAAIYAKGQGIRHIIAGVCETDFSGYPDCRDVFVKSMNVTLNLAMDYDFQIHTPL
                MYLTKAQTWALADEMGALDYIREQTHTCYNGIVGGCRECPSCILRERGLAEYLESKKA
                V"
misc_feature    3522. .3531
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
RBS             3547. .3550
gene            3558. .4076
                /gene="NMA0703"
CDS             3558. .4076
                /gene="NMA0703"
                /note="NMA0703, len: 172 aa; unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0703"
                /protein_id="CAB83990.1"
                /db_xref="GI:7379428"
                /db_xref="SPTREMBL:Q9JVT7"
                /translation="MPKLHMFYLGGNAGRSNIEVHDIQFAVCDDYREAVPALKAAWFG
                DADKIHIDGWQVVEWADGYDIAVSETPKTKMPSENAPRLYFANVGGYRAGQLAEAHAF
                GLFAAATPAEAKQKALQTLLTDSYVQQHKDNLKDVDNLLALDRIGNFHIRLTPNPHGK
                PAEIGFQGYLPI"
misc_feature    3782. .3791
                /gene="NMA0703"
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            4083. .4505
                /gene="NMA0704"
misc_feature    4083. .4502
                /gene="NMA0704"
                /note="Pfam match to entry PF01242 PTPS, 6-pyruvoyl
                tetrahydropterin synthase, score 10.80, E-value 1.7e-06"
CDS             4083. .4505
                /gene="NMA0704"
                /note="NMA0704, len: 140 aa; similar to many hypothetical
                proteins e.g. SW:YB90_HAEIN (EMBL:U32798), HI1190,
                Haemophilus influenzae hypothetical protein (141 aa),
                fasta scores; E(): 5.1e-32, 58.0% identity in 138 aa
                overlap. Shows very weak similarity to eukaryotic
                6-pyruvoyl-tetrahydropterin synthases e.g. SW:PTPS_RAT
                (EMBL:M77850), pts, Rattus norvegicus 6-pyruvoyl
                tetrahydrobiopterin synthase precursor (EC 4.6.1.10) (144
                aa), fasta scores; E(): 0.0062, 27.3% identity in 154 aa
                overlap. Contains Pfam match to entry PF01242 PTPS,
                6-pyruvoyl tetrahydropterin synthase"
```

```
Query Match             65.2%;  Score 1231.2;  DB 1;  Length 311321;
Best Local Similarity   96.3%;  Pred. No. 5.5e-281;
Matches 1260;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy       1 ttgggcatttcccgcaaaatatcccttattctgtccatactggcagtgtgcctgccgatg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142661 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATG 142720

Qy      61 catgcacacgcctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcag 120
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  142721 CATGCACACGCCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAG 142780

Qy     121 catttcgaacccgacgggaaataccacctattcggcagcagggggggagcttgccnagcgc 180
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||| |||||
Db  142781 CATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGC 142840

Qy     181 aacggccatatcggattgggaaacatacaaagccatcagttgggccacctgatgattcaa 240
           | ||| ||||||||||||||||||||||||||||||||||||||||| ||||| | || ||
Db  142841 AGCGGTCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCAACCTGTTCATCCAG 142900

Qy     241 caggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaa 300
           ||||||||| || ||||||||||||||||||||||||||||||||||||||||||||| ||
Db  142901 CAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAA 142960

Qy     301 ttccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtccc 360
           ||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  142961 GTCCATTCCCCCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCC 143020

Qy     361 gttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccgac 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143021 GTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGAC 143080

Qy     421 ggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatac 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143081 GGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATAC 143140

Qy     481 agctacgacataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143141 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGC 143200

Qy     541 accggacaacggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagta 600
           |||||||||||||||| |||||||||||||||| ||||   ||||||||||||||||||
Db  143201 ACCGGACAACGGCTTGTCGACCGTTTCCACAATACCGGTAGTATGCTGACGCAAGGAGTA 143260

Qy     601 ggcgacggattcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgcc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143261 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 143320

Qy     661 gccgaagccttcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaa 720
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143321 GCCGAAGCTTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAA 143380

Qy     721 attgtcggcgcaggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatg 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  143381 ATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTTATG 143440

Qy     781 cacggcttgggtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagat 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143441 CACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 143500

Qy     841 atggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaacccc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143501 ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 143560

Qy     901 aatgccgcacaaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaa 960
           |||||||||||||||||||||||||||||||||||| ||||| ||||||| ||||||
Db  143561 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTACGGCAGTCATCCCCGTCAAA 143620

Qy     961 gggattggagctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaag 1020
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 143621 GGGATTGGAGCTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAG 143680

Qy   1021 cggtcgcagatgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaatttt 1080
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db 143681 CGGTCGCAGATGGGCGAGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTT 143740

Qy   1081 gccgatgcggcatacgccaaatacccgtcccctaccattcccgaaatatccgttcaaac 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143741 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAAC 143800

Qy   1141 ttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143801 TTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGA 143860

Qy   1201 aaaaatgtcaaactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaa 1260
          || ||||| |||||||||| || ||||||||||||||||    || ||||||||||||||||
Db 143861 AAGAATGTGAAACTGGCAAACAAACGCCACCCGAAGACCAAAGTGCCGTTTGACGGTAAA 143920

Qy   1261 gggtttccgaattttgagaagcacgtgaaatatgatacgaagctcgata 1309
          |||||||||||||||| ||  |||| ||||| |||||||   |  |||
Db 143921 GGGTTTCCGAATTTTGAAAAAGACGTAAAATACGATACGAGAATTAATA 143969


RESULT   7
AX236465
LOCUS       AX236465               2304 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 158 from Patent WO0164922.
ACCESSION   AX236465
VERSION     AX236465.1  GI:15796054
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2304)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 158 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2304
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961cL-ORF46.1"
BASE COUNT      710 a    619 c    555 g    420 t
ORIGIN


  Query Match           63.5%;  Score 1198.8;  DB 6;  Length 2304;
  Best Local Similarity  98.1%;  Pred. No. 1.3e-273;
  Matches 1212;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db   1069 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1128

Qy    133 gacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
          ||||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db   1129 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1188

Qy    193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
          |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||| ||
Db   1189 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1248

Qy    253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
          |||||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db   1249 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1308

Qy    313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1309 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1368

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1369 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1428

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1429 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1488

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1489 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1548

Qy    553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
          |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db   1549 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1608

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1609 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1668

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1669 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1728

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1729 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1788

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1789 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1848

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1849 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 1908

Qy    913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1909 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 1968

Qy    973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          || ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1969 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 2028

Qy   1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2029 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2088

Qy   1093 tacgccaaatacccgtcccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2089 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2148

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2149 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2208

Qy   1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2209 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2268

Qy   1273 tttgagaagcacgtgaaatatgatacgaagctcga 1307
          ||||||||||||||||||||||||||| | |||||
Db   2269 TTTGAGAAGCACGTGAAATATGATACGTAACTCGA 2303


RESULT   8
AX239751
LOCUS       AX239751               2304 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 53 from Patent WO0164920.
ACCESSION   AX239751
VERSION     AX239751.1  GI:15797386
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
```

```
          artificial sequence.
REFERENCE   1  (bases 1 to 2304)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
          and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 53 07-SEP-2001;
          Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2304
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961cL-ORF46.1"
BASE COUNT      710 a    619 c    555 g    420 t
ORIGIN


  Query Match           63.5%;  Score 1198.8;  DB 6;  Length 2304;
  Best Local Similarity  98.1%;  Pred. No. 1.3e-273;
  Matches 1212;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db    1069 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1128

Qy     133 gacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
           |||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db    1129 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1188

Qy     193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
           |||||||||||| |||||||||||||||||||| |||||||||||||||||||||||| ||
Db    1189 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1248

Qy     253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
           |||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db    1249 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1308

Qy     313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    1309 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1368

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1369 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1428

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1429 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1488

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1489 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1548

Qy     553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
           |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db    1549 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1608

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1609 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1668

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    1669 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1728

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    1729 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1788

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1789 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1848

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1849 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 1908

Qy     913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1909 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 1968

Qy     973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           || ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1969 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 2028

Qy    1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2029 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2088

Qy    1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2089 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2148

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2149 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2208

Qy    1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2209 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2268

Qy    1273 tttgagaagcacgtgaaatatgatacgaagctcga 1307
           ||||||||||||||||||||||||||| | |||||
Db    2269 TTTGAGAAGCACGTGAAATATGATACGTAACTCGA 2303


RESULT   9
AX236421
LOCUS       AX236421                2028 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 114 from Patent WO0164922.
ACCESSION   AX236421
VERSION     AX236421.1  GI:15796033
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
          artificial sequence.
REFERENCE   1  (bases 1 to 2028)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
          Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 114 07-SEP-2001;
          Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2028
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="deltaG741-ORF46.1"
BASE COUNT      557 a    577 c    531 g    363 t
ORIGIN


  Query Match           63.4%;  Score 1196;  DB 6;  Length 2028;
  Best Local Similarity  98.3%;  Pred. No. 6.1e-273;
  Matches 1208;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      71 cctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaac 130
           |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db     773 CCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 832

Qy     131 ccgacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccata 190
           |||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db     833 CCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATA 892

Qy     191 tcggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccg 250
           |||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||
Db     893 TCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCA 952
```

```
Qy    251 ttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgc 310
          || ||||||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |
Db    953 TTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCC 1012

Qy    311 ccttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggat 370
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1013 CCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGAT 1072

Qy    371 tcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacg 430
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1073 TTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACG 1132

Qy    431 ggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgaca 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 GGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACA 1192

Qy    491 taaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaac 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 TAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 1252

Qy    551 ggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggat 610
          |||||||||||||||||||||||||||    ||||||||||||||||||||||||||||||
Db   1253 GGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGAT 1312

Qy    611 tcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagcct 670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1313 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCT 1372

Qy    671 tcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcg 730
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1373 TCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 1432

Qy    731 caggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgg 790
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1433 CAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 1492

Qy    791 gtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaac 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1493 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 1552

Qy    851 tcaaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcac 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1553 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 1612

Qy    911 aaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggag 970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1613 AAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 1672

Qy    971 ctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcaga 1030
          |||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1673 CTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGA 1732

Qy   1031 tgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcgg 1090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1733 TGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG 1792

Qy   1091 catacgccaaataccgtccccttaccattcccgaaatatccgttcaaacttggagcagc 1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1793 CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 1852

Qy   1151 gttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtca 1210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1853 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCA 1912

Qy   1211 aactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccga 1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1913 AACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 1972

Qy   1271 attttgagaagcacgtgaaatatgatacg 1299
          ||||||||||||||||||||||||||||
Db   1973 ATTTTGAGAAGCACGTGAAATATGATACG 2001
```

```
RESULT  10
AX239731
LOCUS       AX239731               2028 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 33 from Patent WO0164920.
ACCESSION   AX239731
VERSION     AX239731.1  GI:15797376
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2028)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 33 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2028
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="deltaG741-ORF46.1"
BASE COUNT      557 a    577 c    531 g    363 t
ORIGIN
```

Query Match 63.4%; Score 1196; DB 6; Length 2028;
Best Local Similarity 98.3%; Pred. No. 6.1e-273;
Matches 1208; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy     71 cctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaac 130
          |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db    773 CCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 832

Qy    131 ccgacgggaaataccacctattcggcagcagggggggagcttgccnagcgcaacggccata 190
          |||||||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||
Db    833 CCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATA 892

Qy    191 tcggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccg 250
          |||||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||
Db    893 TCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCA 952

Qy    251 ttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgc 310
          || ||||||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |
Db    953 TTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCC 1012

Qy    311 ccttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggat 370
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1013 CCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGAT 1072

Qy    371 tcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacg 430
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1073 TTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACG 1132

Qy    431 ggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgaca 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 GGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACA 1192

Qy    491 taaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaac 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 TAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 1252

Qy    551 ggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggat 610
          |||||||||||||||||||||||||||    ||||||||||||||||||||||||||||||
Db   1253 GGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGAT 1312

Qy    611 tcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagcct 670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1313 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCT 1372

Qy    671 tcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcg 730
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1373 TCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 1432
```

```
Qy    731 caggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgg 790
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   1433 CAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 1492

Qy    791 gtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaac 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1493 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 1552

Qy    851 tcaaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcac 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1553 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 1612

Qy    911 aaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggag 970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1613 AAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 1672

Qy    971 ctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcaga 1030
          |||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1673 CTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGA 1732

Qy   1031 tgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcgg 1090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1733 TGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG 1792

Qy   1091 catacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagc 1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1793 CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 1852

Qy   1151 gttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtca 1210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1853 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCA 1912

Qy   1211 aactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccga 1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1913 AACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 1972

Qy   1271 attttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||||
Db   1973 ATTTTGAGAAGCACGTGAAATATGATACG 2001


RESULT  11
AX236407
LOCUS       AX236407                4425 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 100 from Patent WO0164922.
ACCESSION   AX236407
VERSION     AX236407.1  GI:15796026
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 4425)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 100 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .4425
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="deltaG983-ORF46.1"
BASE COUNT     1165 a   1259 c   1219 g    782 t
ORIGIN


  Query Match             63.4%;  Score 1196;  DB 6;  Length 4425;
  Best Local Similarity   98.3%;  Pred. No. 6.8e-273;
  Matches 1208;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy     71 cctcagatttggcaaacgatcccttatccggcaggttctcgaccgtcagcatttcgaac 130
          |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db   3170 CCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 3229

Qy    131 ccgacgggaaaataccacctattcggcagcagggggagcttgccnagcgcaacggccata 190
          |||||||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||
Db   3230 CCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATA 3289

Qy    191 tcggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccg 250
          |||||||||||| |||||||||||||||||||| ||||||||||||||||||||||||||
Db   3290 TCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCA 3349

Qy    251 ttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgc 310
          || |||||||||||||||||||||||||||||||||||||||||||| || ||||||| |
Db   3350 TTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCC 3409

Qy    311 ccttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggat 370
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   3410 CCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGAT 3469

Qy    371 tcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacg 430
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3470 TTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACG 3529

Qy    431 ggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgaca 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3530 GGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACA 3589

Qy    491 taaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaac 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3590 TAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 3649

Qy    551 ggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggat 610
          |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db   3650 GGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGAT 3709

Qy    611 tcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagcct 670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3710 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCT 3769

Qy    671 tcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcg 730
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3770 TCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 3829

Qy    731 caggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgg 790
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   3830 CAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 3889

Qy    791 gtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaac 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3890 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 3949

Qy    851 tcaaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcac 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3950 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 4009

Qy    911 aaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggag 970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4010 AAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 4069

Qy    971 ctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcaga 1030
          |||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   4070 CTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGA 4129

Qy   1031 tgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcgg 1090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4130 TGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG 4189

Qy   1091 catacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagc 1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4190 CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 4249

Qy   1151 gttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtca 1210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4250 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCA 4309
```

```
Qy   1211 aactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccga 1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4310 AACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 4369

Qy   1271 attttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||||
Db   4370 ATTTTGAGAAGCACGTGAAATATGATACG 4398


RESULT 12
AX239715
LOCUS       AX239715               4425 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 17 from Patent WO0164920.
ACCESSION   AX239715
VERSION     AX239715.1  GI:15797369
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 4425)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 17 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .4425
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="deltaG983-ORF46.1"
BASE COUNT     1165 a   1259 c   1219 g    782 t
ORIGIN


  Query Match             63.4%;  Score 1196;  DB 6;  Length 4425;
  Best Local Similarity   98.3%;  Pred. No. 6.8e-273;
  Matches 1208;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy     71 cctcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaac 130
          |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db   3170 CCTCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 3229

Qy    131 ccgacgggaaataccacctattcggcagcagggggagcttgccnagcgcaaccggccata 190
          |||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||
Db   3230 CCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATA 3289

Qy    191 tcggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccg 250
          |||||||||||||| |||||||||||||||||||| |||||||||||||||||||||||
Db   3290 TCGGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCA 3349

Qy    251 ttgaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgc 310
          || |||||||||||||||||||||||||||||||||||||||||||||| || |||||||| |
Db   3350 TTAAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCC 3409

Qy    311 ccttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggat 370
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   3410 CCTTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGAT 3469

Qy    371 tcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacg 430
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3470 TTAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACG 3529

Qy    431 ggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgaca 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3530 GGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACA 3589

Qy    491 taaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaac 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3590 TAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 3649

Qy    551 ggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggat 610
          |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db   3650 GGCTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGAT 3709

Qy    611 tcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagcct 670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3710 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCT 3769

Qy    671 tcaacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcg 730
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   3770 TCAACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 3829

Qy    731 caggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgg 790
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3830 CAGGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 3889

Qy    791 gtctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaac 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3890 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 3949

Qy    851 tcaaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcac 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3950 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 4009

Qy    911 aaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggag 970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4010 AAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 4069

Qy    971 ctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcaga 1030
          |||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   4070 CTGTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGA 4129

Qy   1031 tgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcgg 1090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4130 TGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG 4189

Qy   1091 catacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagc 1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4190 CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 4249

Qy   1151 gttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtca 1210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4250 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCA 4309

Qy   1211 aactggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccga 1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4310 AACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 4369

Qy   1271 attttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||||
Db   4370 ATTTTGAGAAGCACGTGAAATATGATACG 4398


RESULT 13
AX236447
LOCUS       AX236447               2019 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 140 from Patent WO0164922.
ACCESSION   AX236447
VERSION     AX236447.1  GI:15796045
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2019)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 140 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2019
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-741"
BASE COUNT      555 a    571 c    531 g    362 t
ORIGIN
```

```
  Query Match             63.3%;  Score 1194;  DB 6;  Length 2019;
  Best Local Similarity   98.3%;  Pred. No. 1.8e-272;
  Matches 1206;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db       4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy     133 gacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
           ||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db      64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy     193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
           |||||||||||| |||||||||||||||||||| |||||||||||||||||||||||| ||
Db     124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy     253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
           ||||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db     184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy     313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy     553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
           |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db     484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy     913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy     973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           || ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy    1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy    1093 tacgccaaatacccgtcccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy    1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy    1273 tttgagaagcacgtgaaatatgatacg 1299
           |||||||||||||||||||||||||||
Db    1204 TTTGAGAAGCACGTGAAATATGATACG 1230
```

```
RESULT  14
AX239733
LOCUS       AX239733               2019 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 35 from Patent WO0164920.
ACCESSION   AX239733
VERSION     AX239733.1  GI:15797377
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2019)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 35 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2019
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-741"
BASE COUNT      555 a    571 c    531 g    362 t
ORIGIN
```

```
  Query Match             63.3%;  Score 1194;  DB 6;  Length 2019;
  Best Local Similarity   98.3%;  Pred. No. 1.8e-272;
  Matches 1206;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db       4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy     133 gacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
           ||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db      64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy     193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
           |||||||||||| |||||||||||||||||||| |||||||||||||||||||||||| ||
Db     124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy     253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
           ||||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db     184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy     313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy     553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
           |||||||||||||||||||||||||    |||||||||||||||||||||||||||||||
Db     484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db     604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
           ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db     664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy     913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy     973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           || ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy    1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy    1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy    1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy    1273 tttgagaagcacgtgaaatatgatacg 1299
           |||||||||||||||||||||||||||
Db    1204 TTTGAGAAGCACGTGAAATATGATACG 1230


RESULT  15
AX236451
LOCUS       AX236451               2256 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 144 from Patent WO0164922.
ACCESSION   AX236451
VERSION     AX236451.1  GI:15796047
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 144 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-961c"
BASE COUNT      698 a    608 c    545 g    405 t
ORIGIN


  Query Match           63.3%;  Score 1194;  DB 6;  Length 2256;
  Best Local Similarity 98.3%;  Pred. No. 1.9e-272;
  Matches 1206;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db       4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy     133 gacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
           |||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db      64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy     193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
           |||||||||| |||||||||||||||||||| |||||||||||||||||||||||| ||
Db     124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy     253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
           |||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db     184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy     313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy     553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
           |||||||||||||||||||||||||    |||||||||||||||||||||||||||||||
Db     484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db     604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
           ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db     664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy     913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy    973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          || |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy   1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy   1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy   1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy   1273 tttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||
Db   1204 TTTGAGAAGCACGTGAAATATGATACG 1230


RESULT 16
AX236459
LOCUS       AX236459                2256 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 152 from Patent WO0164922.
ACCESSION   AX236459
VERSION     AX236459.1  GI:15796051
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 152 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961c-ORF46.1"
BASE COUNT      698 a    608 c    545 g    405 t
ORIGIN


  Query Match             63.3%;  Score 1194;  DB 6;  Length 2256;
  Best Local Similarity   98.3%;  Pred. No. 1.9e-272;
  Matches 1206;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy     73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          |||||||||||||||||| | |||||||||||||||||||||||||||||||||||||||
Db   1003 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1062

Qy    133 gacgggaaatacccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
          |||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db   1063 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1122

Qy    193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
          |||||||||||| |||||||||||||||||||| |||||||||||||||||||||||| ||
Db   1123 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1182

Qy    253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
          |||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db   1183 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1242

Qy    313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1243 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1302

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1303 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1362

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1363 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1422

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1423 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1482

Qy    553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
          |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db   1483 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1542

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1543 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1602

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1603 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1662

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1663 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1722

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1782

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1783 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 1842

Qy    913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1843 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 1902

Qy    973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          || |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1903 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 1962

Qy   1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1963 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2022

Qy   1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2023 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2082

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2083 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2142

Qy   1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2143 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2202

Qy   1273 tttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||
Db   2203 TTTGAGAAGCACGTGAAATATGATACG 2229


RESULT 17
AX239737
LOCUS       AX239737                2256 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 39 from Patent WO0164920.
ACCESSION   AX239737
```

```
VERSION     AX239737.1  GI:15797379
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 39 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-961c"
BASE COUNT      698 a    608 c    545 g    405 t
ORIGIN
```

Query Match 63.3%; Score 1194; DB 6; Length 2256;
Best Local Similarity 98.3%; Pred. No. 1.9e-272;
Matches 1206; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy      73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db       4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy     133 gacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
           |||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db      64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy     193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
           |||||||||| |||||||||||||||||| |||||||||||||||||||||||||| ||
Db     124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy     253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
           |||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db     184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy     313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy     553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
           |||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db     484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy     913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy     973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           || |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy    1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy    1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy    1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy    1273 tttgagaagcacgtgaaatatgatacg 1299
           |||||||||||||||||||||||||||
Db    1204 TTTGAGAAGCACGTGAAATATGATACG 1230
```

```
RESULT  18
AX239745
LOCUS       AX239745                2256 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 47 from Patent WO0164920.
ACCESSION   AX239745
VERSION     AX239745.1  GI:15797383
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 47 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961c-ORF46.1"
BASE COUNT      698 a    608 c    545 g    405 t
ORIGIN
```

Query Match 63.3%; Score 1194; DB 6; Length 2256;
Best Local Similarity 98.3%; Pred. No. 1.9e-272;
Matches 1206; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy      73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db    1003 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1062

Qy     133 gacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
           |||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db    1063 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1122
```

```
Qy    193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
          |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||| ||
Db   1123 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1182

Qy    253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
          |||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db   1183 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1242

Qy    313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1243 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1302

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1303 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1362

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1363 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1422

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1423 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1482

Qy    553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
          |||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db   1483 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1542

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1543 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1602

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1603 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1662

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1663 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1722

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1782

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1783 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 1842

Qy    913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1843 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 1902

Qy    973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          || |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1903 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 1962

Qy   1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1963 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2022

Qy   1093 tacgccaaatacccgtcccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2023 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2082

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2083 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2142

Qy   1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2143 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2202

Qy   1273 tttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||
Db   2203 TTTGAGAAGCACGTGAAATATGATACG 2229
```

```
RESULT 19
AX236449
LOCUS       AX236449               2421 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 142 from Patent WO0164922.
ACCESSION   AX236449
VERSION     AX236449.1  GI:15796046
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2421)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 142 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2421
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-961"
BASE COUNT      730 a    659 c    591 g    441 t
ORIGIN
```

Query Match 63.3%; Score 1194; DB 6; Length 2421;
Best Local Similarity 98.3%; Pred. No. 1.9e-272;
Matches 1206; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy     73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db      4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy    133 gacgggaaataccacctattcggcagcaggggggagcttgccnagcgcaacggccatatc 192
          |||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db     64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy    193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
          |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||| ||
Db    124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy    253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
          |||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db    184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy    313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy    553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
          |||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db    484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603
```

```
Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy    913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy    973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          || ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy   1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy   1093 tacgccaaatacccgtcccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy   1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy   1273 tttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||
Db   1204 TTTGAGAAGCACGTGAAATATGATACG 1230
```

```
RESULT  20
AX236453
LOCUS       AX236453                2421 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 146 from Patent WO0164922.
ACCESSION   AX236453
VERSION     AX236453.1  GI:15796048
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2421)
  AUTHORS   Arico,M.B., Comanducci,M.C., Galeotti,C.C., Masignani,V.C.,
            Guiliani,M.M. and Pizza,M.C.
  TITLE     Heterologous expression of neisserial proteins
  JOURNAL   Patent: WO 0164922-A 146 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2421
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="961-ORF46.1"
BASE COUNT      730 a    659 c    591 g    441 t
ORIGIN
```

```
  Query Match           63.3%;  Score 1194;  DB 6;  Length 2421;
  Best Local Similarity  98.3%;  Pred. No. 1.9e-272;
  Matches 1206;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;
```

```
Qy     73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db   1168 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1227

Qy    133 gacgggaaataccacctattcggcagcagggggggagcttgccnagcgcaacggccatatc 192
          |||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db   1228 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1287

Qy    193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
          |||||||||| |||||||||||||||||||| |||||||||||||||||||||||| ||
Db   1288 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1347

Qy    253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
          |||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db   1348 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1407

Qy    313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1408 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1467

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1468 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1527

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1528 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1587

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1588 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1647

Qy    553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
          |||||||||||||||||||||||||    |||||||||||||||||||||||||||||||
Db   1648 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1707

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1708 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1767

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   1768 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1827

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1828 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1887

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1888 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1947

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1948 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 2007

Qy    913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2008 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 2067

Qy    973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          || ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   2068 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 2127

Qy   1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2128 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2187

Qy   1093 tacgccaaatacccgtcccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2188 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2247
```

```
Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2248 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2307

Qy   1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2308 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2367

Qy   1273 tttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||
Db   2368 TTTGAGAAGCACGTGAAATATGATACG 2394


RESULT  21
AX239735
LOCUS       AX239735               2421 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 37 from Patent WO0164920.
ACCESSION   AX239735
VERSION     AX239735.1  GI:15797378
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2421)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 37 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2421
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="ORF46.1-961"
BASE COUNT      730 a    659 c    591 g    441 t
ORIGIN


  Query Match             63.3%;  Score 1194;  DB 6;  Length 2421;
  Best Local Similarity   98.3%;  Pred. No. 1.9e-272;
  Matches 1206;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy     73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
          |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db      4 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 63

Qy    133 gacgggaaataccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
          |||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db     64 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 123

Qy    193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
          |||||||||||| |||||||||||||||||||| |||||||||||||||||||||| ||
Db    124 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 183

Qy    253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
           |||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db    184 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 243

Qy    313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    244 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 303

Qy    373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 363

Qy    433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 423

Qy    493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 483

Qy    553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
          |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db    484 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 543

Qy    613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 603

Qy    673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    604 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 663

Qy    733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    664 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 723

Qy    793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    724 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 783

Qy    853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    784 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 843

Qy    913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    844 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 903

Qy    973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
          || ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    904 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 963

Qy   1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 1023

Qy   1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1083

Qy   1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1084 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 1143

Qy   1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1203

Qy   1273 tttgagaagcacgtgaaatatgatacg 1299
          |||||||||||||||||||||||||||
Db   1204 TTTGAGAAGCACGTGAAATATGATACG 1230


RESULT  22
AX239739
LOCUS       AX239739               2421 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 41 from Patent WO0164920.
ACCESSION   AX239739
VERSION     AX239739.1  GI:15797380
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 2421)
  AUTHORS   Arico,M.B., Comanducci,M., Galeotti,C., Masignani,V., Giuliani,M.M.
            and Pizza,M.
  TITLE     Hybrid expression of neisserial proteins
  JOURNAL   Patent: WO 0164920-A 41 07-SEP-2001;
            Chiron Spa (IT)
FEATURES             Location/Qualifiers
     source          1. .2421
                     /organism="synthetic construct"
```

```
                     /db_xref="taxon:32630"
                     /note="961-ORF46.1"
BASE COUNT       730 a    659 c    591 g    441 t
ORIGIN

  Query Match           63.3%;  Score 1194;  DB 6;  Length 2421;
  Best Local Similarity  98.3%;  Pred. No. 1.9e-272;
  Matches 1206;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      73 tcagatttggcaaacgatccctttatccggcaggttctcgaccgtcagcatttcgaaccc 132
           |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db    1168 TCAGATTTGGCAAACGATTCTTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 1227

Qy     133 gacgggaaatacccacctattcggcagcagggggagcttgccnagcgcaacggccatatc 192
           ||||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||||
Db    1228 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAACTTGCCGAGCGCAGCGGCCATATC 1287

Qy     193 ggattgggaaacatacaaagccatcagttgggccacctgatgattcaacaggcggccgtt 252
           |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||| ||
Db    1288 GGATTGGGAAAAATACAAAGCCATCAGTTGGGCAACCTGATGATTCAACAGGCGGCCATT 1347

Qy     253 gaaggaaatatcggctacattgtccgcttttccgatcacgggcacaaattccattcgccc 312
           |||||||||||||||||||||||||||||||||||||||||||||||| || ||||||| |||
Db    1348 AAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACGAAGTCCATTCCCCC 1407

Qy     313 ttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattc 372
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1408 TTCGACAACCATGCCTCACATTCCGATTCTGATGAAGCCGGTAGTCCCGTTGACGGATTT 1467

Qy     373 agcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacggg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1468 AGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 1527

Qy     433 ccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacata 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1528 CCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 1587

Qy     493 aaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacgg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1588 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGG 1647

Qy     553 cttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattc 612
           |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db    1648 CTTGCCGACCGTTTCCACAATGCCGGTAGTATGCTGACGCAAGGAGTAGGCGACGGATTC 1707

Qy     613 aaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1708 AAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 1767

Qy     673 aacggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgca 732
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1768 AACGGCACTGCAGATATCGTTAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 1827

Qy     733 ggcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatgcacggcttgggt 792
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1828 GGCGATGCCGTGCAGGGCATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 1887

Qy     793 ctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaactc 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1888 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 1947

Qy     853 aaagactatgccgcagcagccatccgcgattgggcagtccaaaaccccaatgccgcacaa 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1948 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAA 2007

Qy     913 ggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattggagct 972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2008 GGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 2067

Qy     973 gtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgcagatg 1032
           || |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    2068 GTTCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTATCAAGCGGTCGCAGATG 2127

Qy    1033 ggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatgcggca 1092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2128 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCA 2187

Qy    1093 tacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagcagcgt 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2188 TACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 2247

Qy    1153 tacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatgtcaaa 1212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2248 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA 2307

Qy    1213 ctggcagaccaacgccacccgaagacaggcgtaccgtttgacggtaaagggtttccgaat 1272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2308 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 2367

Qy    1273 tttgagaagcacgtgaaatatgatacg 1299
           |||||||||||||||||||||||||||
Db    2368 TTTGAGAAGCACGTGAAATATGATACG 2394


RESULT  23
A96426
LOCUS       A96426                   683 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 459 from Patent WO9924578.
ACCESSION   A96426
VERSION     A96426.1  GI:6780105
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 459 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      176 a    202 c    171 g    131 t      3 others
ORIGIN

  Query Match           34.1%;  Score 642.8;  DB 6;  Length 683;
  Best Local Similarity  96.0%;  Pred. No. 8.8e-142;
  Matches  656;  Conservative    1;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      43 gcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgatccctttatccgg 102
           |||||||||| |||||||||||||||||||| ||||||||||||| | |||||||||
Db       1 GCAGTGTGCCTNCCGATGCATGCACACGCCTCANATTTGGCAAACGATTCTTTTATCCGG 60

Qy     103 caggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcggcagcagg 162
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGG 120

Qy     163 ggggagcttgccnagcgcaacggccatatcggattgggaaacatacaaagccatcagttg 222
           ||||| |||||| |||| |    ||||||||||||||| ||||||||||||||||||||
Db     121 GGGGAACTTGCCGAGCGCCAGTCTCATATCGGATTGGGAAAAATACAAAGCCATCAGTTG 180

Qy     223 ggccacctgatgattcaacaggcggccgttgaaggaaatatcggctacattgtccgcttt 282
           ||| |||||||||||||||||||| || ||||||||||||||||||||||||||||||||
Db     181 GGCAACCTGATGATTCAACAGGCGGCCATTAAAGGAAATATCGGCTACATTGTCCGCTTT 240

Qy     283 tccgatcacgggcacaaattccattcgccttcgacaaccatgcctcacattccgattct 342
           |||||||||||||| || ||||||| ||:|||||||||||||||||||||||||||||||
Db     241 TCCGATCACGGGCACGAAGTCCATTCCCCSTTCGACAACCATGCCTCACATTCCGATTCT 300

Qy     343 gacgaagccggtagtcccgttgacggattcagcctttaccgcatccattgggacggatac 402
           || |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     301 GATGAAGCCGGTAGTCCCGTTGACGGATTTAGCCTTTACCGCATCCATTGGGACGGATAC 360
```

```
Qy    403 gaacaccatcccgccgacggctatgacgggccacagggcggcggctatcccgctcccaaa 462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAA 420

Qy    463 ggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaatatccgcctcaac 522
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    421 GGCGCGAGGGATATATACAGTTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAAC 480

Qy    523 ctgaccgacaaccgcagcaccggacaacggcttgccgaccgtttccacaatgccggcgct 582
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||    |
Db    481 CTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGTAGT 540

Qy    583 atgctgacgcaaggagtaggcgacggattcaaacgcgccacccgatacagccccgagctg 642
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTG 600

Qy    643 gacagatcgggcaatgccgccgaagccttcaacggcactgcagatatcgtcaaaaacatc 702
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    601 GACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTTAAAAACATC 660

Qy    703 atcggcgcggcaggagaaattgt 725
          |||||||| ||||||||||||||
Db    661 ATCGGCGCTGCAGGAGAAATTGT 683


RESULT  24
A96424
LOCUS       A96424                     381 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 457 from Patent WO9924578.
ACCESSION   A96424
VERSION     A96424.1  GI:6780104
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 381)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 457 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .381
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      127 a     54 c    108 g     90 t      2 others
ORIGIN


  Query Match           17.1%;  Score 322.4;  DB 6;  Length 381;
  Best Local Similarity  99.1%;  Pred. No. 8.1e-66;
  Matches  334;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy   1320 atcgggggggcggtatacctaaggctaagcctgtgtttgatgcgaaaccgagatgggaggt 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 ATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGT 104

Qy   1380 tgataggaagcttaataaattgacaactcgtgagcaggtggagaaaaatgttcaggaaac 1439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 TGATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAAC 164

Qy   1440 gagaagaaggagtcagagtagtcagtttaaagcccatgcgcaacgagaatgggaaaataa 1499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 GAGAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAAATAA 224

Qy   1500 aacagggttagatttttaatcatttttataggtggtgatatcaataagaaaggcacagtaac 1559
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    225 AACAGGGTTAGATTTTAATCATTTTATAGGTGGTGATATCAATAAAAAAGGCACAGTAAC 284

Qy   1560 aggagggcatagtctaacccgtggtgatgtacgggtgatacaacaaacctcggcacctga 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 AGGAGGGCATAGTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTGA 344

Qy   1620 taaacatggggt-ttatcaagcgacagtggaaaattaa 1655
          |||||||||||| |||||||||||||||||||||| |
Db    345 TAAACATGGGGTNTTATCAAGCGACAGTGGAAATTNA 381


RESULT  25
AE002393/c
LOCUS       AE002393                11700 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 35 of 206 of
            the complete genome.
ACCESSION   AE002393 AE002098
VERSION     AE002393.1  GI:7225586
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 11700)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 11700)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11700
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            201. .974
                     /gene="NMB0364"
     CDS             201. .974
                     /gene="NMB0364"
                     /note="similar to GB:L06299 PID:293962 percent identity:
                     99.61; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="FrpC operon protein"
                     /protein_id="AAF40807.1"
                     /db_xref="GI:7225587"
                     /translation="MRPYATTIYQLFILFIGSVFTMTSCEPVNEQTSFNNPEPMTGFE
                     HTVTFDFQGTKMVIPYGYLARYTQNNATKWLSDTPGQDAYSINLIEISVYYKKTDQGW
                     VLEPYNQQNKAHFIQFLRDGLDSVDDIVIRKDACSLSTTMGERLLTYGVKKMPSAYPE
                     YEAYEDKRHIPENPYFHEFYYIKKGENPAIITHRNNRINQTEEDSYSTSVGSCINGFT
                     VRYYPFIREKQQLTQQELVGYHQQVEQLVQSFVNNSSKK"
     gene            991. .1266
                     /gene="NMB0365"
                     /note="This region contains a match to at least one other
                     gene that is not full length,and is not the result of a
                     sequencing artifact; similar to GB:L06299 PID:293963
```

SP:P55127 percent identity: 98.72; identified by sequence similarity; putative"

gene complement(1353. .2486)
/gene="NMB0366"

CDS complement(1353. .2486)
/gene="NMB0366"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF40808.1"
/db_xref="GI:7225588"
/translation="MFKKFKPVLLSFFALVFAFWLGTGIAYEINPRWFLSDTATEVPK
NPNAFVAKLARLFRNADRAVVIVKESIRTEENLAGTVDDGPLQSEKDYLALAIRLSRL
KEKAKWFHVTEQEHGKEVWLDYHIGEGGLVAVSLSQRSPEAFVNAEYLYRNDRPFSVN
VYGGTVHGENYETTGEYRVVWQPDGSVFDAAGRGKIGEDVYEHCLGCYQMAQVYLAKY
RDVANDEQKVWDFRKESNRIASDSRNSVFYQNMRELMPRGMKANSLVVGYDADGLPQK
VYWSFDNGKKRQSFEYYLKNGNLFIAQSSTVALKADGVTADMQTYHAQQTWYLDGGRI
VREEKQGDRLPDFPLNLENLEKEVRRYAEAAARRSGGRRDLSH"

gene complement(2550. .2936)
/gene="NMB0367"

CDS complement(2550. .2936)
/gene="NMB0367"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF40809.1"
/db_xref="GI:7225589"
/translation="MCIMCELKNFRRNITCFEGYDENSFIGKWYDDGVWDDEEYWKLE
NDLIEVRKKYPYPMDIPRYVVIGIGTIIDFLMVPNWKLFEIKASPWLPDSVGINERYE
RLKTMLRYIFTEKDIVNVQFDYYNKK"

gene complement(3005. .3463)
/gene="NMB0368"

CDS complement(3005. .3463)
/gene="NMB0368"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF40810.1"
/db_xref="GI:7225590"
/translation="MVDAVVKTPEFLPFTSVGIFRFGADITQYKNILETFMYEPPDEF
RTEYYESPDSNLLISVKKKNKIISIFCYQELYFMGVNIIGLNFEDFKQLFHYPKSYGV
DKYYLSNESYPTYVYEFDEIGVQAWEAKGKIVTIIAGGKDNYSTEPYYDE"

gene complement(3444. .3833)
/gene="NMB0369"

CDS complement(3444. .3833)
/gene="NMB0369"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF40811.1"
/db_xref="GI:7225591"
/translation="MKAGSRLLSESGRTQTVRKTVVKPKPLKAYNLTVADWHTYFVKG
NQAETEGVWVHNSCPPKRTGSSKNEKHGDGGRSQISAESKIAELTNKIIPGMSKNERL
KIKQKIRNIAKNANRKTKGEEHGRRGR"

gene complement(3965. .4342)
/gene="NMB0370"

CDS complement(3965. .4342)
/gene="NMB0370"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF40812.1"
/db_xref="GI:7225592"
/translation="MCEFKDIIRNVPYFEGYDENSFIGKWYDDGVWDDEEYWKLENDL
IEVRKKYPYPMDIPRYVVIGIGTIIDFLMVPNWKLFEIKASPWLPDSVGIHERYERFT
TMLRYIFTEKDIVNVRFDYYNKK"

gene complement(4364. .4771)
/gene="NMB0371"

CDS complement(4364. .4771)
/gene="NMB0371"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF40813.1"
/db_xref="GI:7225593"
/translation="MCEFKDFRRNIPCFKEYDENSFIGKWYDDGVWDDEEYWKLENDL
IEVRKKYPYPMDIPRDIVIGIGTIIDFLMVQNWKLFEIKASTWLPKSVKINERYERFR
VMLRYIFTDLDVEDWKFFYFPIQHSKGRLRRKV"

gene complement(4777. .5454)
/gene="NMB0372"

CDS complement(4777. .5454)
/gene="NMB0372"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF40814.1"
/db_xref="GI:7225594"
/translation="MVKTADGYKAIARIRTGDRVFAKDEASGKTGYKPVTARYGNPYQ
ETVYIEISDGIGNNQTLISNKIHPFYSQGKWIQAGRLKKGDTLLSESGAKQTVQNITL
KQQPLKAYNLTVADWHTYFVKGNRAETEGVWVHNDCPPRKTPSTPIYGNDSEAYAAAK
ELGYRKIKERTRNDAAIFKKGKSYISRDVDSHNGGAWKEASSPEKLNRKETRNGTFDK
NLNRIGD"

gene complement(5669. .6028)
/gene="NMB0373"

CDS complement(5669. .6028)
/gene="NMB0373"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF40815.1"
/db_xref="GI:7225595"
/translation="MASYVSIKGWIECSDDDIKIIQENINNFWNNCPFNIEEKESAKI
YKSGWVFPTNSFNWSSYIFFGACVKSYFIIYFEKCIKTIMELDIEISGFFELDYYEDN
YKVNWKINNGNLIQTTN"

gene complement(6031. .7434)
/gene="NMB0374"

CDS complement(6031. .7434)
/gene="NMB0374"
/note="MafB-related protein; identified by Oxford; putative"
/codon_start=1
/transl_table=11
/product="MafB-related protein"
/protein_id="AAF62308.1"
/db_xref="GI:7413426"
/translation="MKPLRRLTNLLAACAVAAAALIQPALAADLAQDPFITDNAQRQH
YEPGGKYHLFGDPRGSVSDRTGKINVIQDYTHQMGNLLIQQANINGTIGYHTRFSGHG
HEEHAPFDNHAADSASEEKGNVDEGFTVYRLNWEGHEHHPADAYDGPKGGNYPKPTGA
RDEYTYHVNGTARSIKLNPTDTRSIRQRISDNYSNLGSNFSDRADEANRKMFEHNAKL

Query Match 15.0%; Score 283.4; DB 1; Length 11700;
Best Local Similarity 57.5%; Pred. No. 2.3e-56;
Matches 528; Conservative 0; Mismatches 387; Indels 3; Gaps 1;

Qy 35 ccatactggcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgatccct 94
|| || ||| | || || | ||| | || ||||| | || || |
Db 7391 CCGTAGCGGCGGCCGCACTCATACAGCCCGCCCTCGCGGCGGACTTGGCGCAAGACCCGT 7332

Qy 95 ttatccggcaggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcg 154
| || | | | || ||||| | |||||||| ||| |||||||||||| ||||

```
Db   7331 TCATTACCGATAACGCCCAACGGCAGCACTACGAACCCGGCGGCAAATACCACCTCTTCG 7272

Qy    155 gcagcagggggagc---ttgccnagcgcaacggccatatcggattgggaaacatacaaa 211
          ||  |  | | |      || || | |||| |||| | |||    |     |     | |
Db   7271 GCGACCCGCGCGGCAGCGTTTCCGACCGCACCGGCAAAATCAACGTCATCCAAGACTATA 7212

Qy    212 gccatcagttgggccacctgatgattcaacaggcggccgttgaaggaaatatcggctaca 271
          ||| ||| ||||| ||||| | || |||||||||    | |  | || |  |||||||||
Db   7211 CCCACCAGATGGGCAACCTGCTCATCCAACAGGCAAACATCAACGGCACAATCGGCTACC 7152

Qy    272 ttgtccgcttttccgatcacgggcacaaattccattcgccccttcgacaaccatgcctcac 331
          ||||||||||  ||||| ||| ||   ||  | ||||||||||||||| ||| |
Db   7151 ACACCCGCTTTTCCGGACACGGACACGAAGAACACGCCCCCTTCGACAACCACGCCGCCG 7092

Qy    332 attccgattctgacgaagccggtagtcccgttgacggattcagcctttaccgcatccatt 391
          |   ||     || |||   || |     |  || || || | | | |||||  || | |
Db   7091 ACAGCGCGAGCGAAGAAAAAGGCAACGTTGACGAAGGCTTTACCGTATACCGGCTCAACT 7032

Qy    392 gggacggatacgaacaccatcccgccgacggctatgacgggccacagggcggcggctatc 451
          |||| ||| ||||||| |||||||||| | ||| ||||| ||  |||||||   || |
Db   7031 GGGAAGGACACGAACATCATCCCGCCGATGCCTACGACGGCCCGAAGGGCGGCAATTACC 6972

Qy    452 ccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaata 511
          ||   || |  |||||  | ||   ||||| |||  || | || |||   ||||  | ||
Db   6971 CCAAACCTACGGGCGCACGAGACGAATACACCTATCACGTCAACGGCACAGCCCGCAGTA 6912

Qy    512 tccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgccgaccgtttccaca 571
          ||   ||||| | |||||||| |||||||| | | |||||  |  |||||  || |  ||
Db   6911 TCAAACTCAATCCGACCGACACCCGCAGCATCCGGCAACGCATATCCGACAATTACAGCA 6852

Qy    572 atgccggcgctatgctgacgcaaggagtaggcgacggattcaaacgcgccacccgataca 631
          |  ||||   |  | |  | | |  || |   || |         |    |||
Db   6851 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 6792

Qy    632 gccccgagctggacagatcgggcaatgccgccgaagccttcaacggcactgcagatatcg 691
            || |||| ||| | | ||||||   |   ||     ||||||||   || |    ||
Db   6791 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 6732

Qy    692 tcaaaaacatcatcggcgcggcaggagaaattgtcggcgcaggcgatgccgtgcagggta 751
            |||  | |   |  || || |||    | |||  ||||||     || | || |
Db   6731 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCTTGGGCATAGGCGACATACTGTACGGAA 6672

Qy    752 taagcgaaggctcaaacattgctgtcatgcacggcttgggtctgctttccaccgaaaaca 811
          || | | |  | |||  || |  |||| |  | | |   |   |  ||  |||   ||
Db   6671 CGCGCTATGCCATAGACAAAGCCGCAATGCGCAACATCGCCCCCTTGCCCGCCGAGGGCA 6612

Qy    812 agatggcgcgcatcaacgatttggcagatatggcgcaactcaaagactatgccgcagcag 871
          |  | ||   ||||  ||  ||||      |||||    |  || |  || |   | ||
Db   6611 AATTCGCCGTCATCGGCGGCTTGGGCAGCGTGGCGGGCTTTGAAAAGAATACGCGCGAAG 6552

Qy    872 ccatccgcgattgggcagtccaaaaccccaatgccgcacaaggcatagaagccgtcagca 931
          || |   |   |||  |    |||| ||||||||||  ||  | | |||||||||  ||
Db   6551 CCGTTGACCGGTGGATACAGGAAAATCCCAATGCCGCCGAAACCGTCGAAGCCGTCTTCA 6492

Qy    932 atatctttatggcagcca 949
          |  |       |||||||
Db   6491 ACGTTGCCGCAGCAGCCA 6474


RESULT  26
AX044029/c
LOCUS       AX044029              349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 108 from Patent WO0066791.
ACCESSION   AX044029
VERSION     AX044029.1  GI:11342913
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 108 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.-seq 1: 1 to
                     349980 349980 bases-seq 108: 300001 to 649980 349980
                     bases-seq 109: 600001 to 949980 349980 bases-seq 110:
                     900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
                     349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
                     113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT     84410 a  84863 c  94187 g  86520 t
ORIGIN


  Query Match             15.0%;  Score 283.4;  DB 6;  Length 349980;
  Best Local Similarity   57.5%;  Pred. No. 3.8e-56;
  Matches  528;  Conservative    0;  Mismatches  387;  Indels    3;  Gaps    1;

Qy     35 ccatactggcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgatccct 94
          || ||  ||| |     ||      ||  || | |||  | || ||||  | || || |
Db  77278 CCGTAGCGGCGGCCGCACTCATACAGCCCGCCCTCGCGGCGGACTTGGCGCAAGACCCGT 77219

Qy     95 ttatccggcaggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcg 154
          | ||     |       | | || |||||| | |||||||| ||| ||||||||||| ||||
Db  77218 TCATTACCGATAACGCCCAACGGCAGCACTACGAACCCGGCGGCAAATACCACCTCTTCG 77159

Qy    155 gcagcagggggagc---ttgccnagcgcaacggccatatcggattgggaaacatacaaa 211
          ||  |  | | |      || || | |||| |||| | |||    |     |     | |
Db  77158 GCGACCCGCGCGGCAGCGTTTCCGACCGCACCGGCAAAATCAACGTCATCCAAGACTATA 77099

Qy    212 gccatcagttgggccacctgatgattcaacaggcggccgttgaaggaaatatcggctaca 271
          ||| ||| ||||| ||||| | || |||||||||    | |  | || |  |||||||||
Db  77098 CCCACCAGATGGGCAACCTGCTCATCCAACAGGCAAACATCAACGGCACAATCGGCTACC 77039

Qy    272 ttgtccgcttttccgatcacgggcacaaattccattcgccccttcgacaaccatgcctcac 331
          ||||||||||  ||||| ||| ||   ||  | ||||||||||||||| ||| |
Db  77038 ACACCCGCTTTTCCGGACACGGACACGAAGAACACGCCCCCTTCGACAACCACGCCGCCG 76979

Qy    332 attccgattctgacgaagccggtagtcccgttgacggattcagcctttaccgcatccatt 391
          |   ||     || |||   || |     |  || || || | | | |||||  || | |
Db  76978 ACAGCGCGAGCGAAGAAAAAGGCAACGTTGACGAAGGCTTTACCGTATACCGGCTCAACT 76919

Qy    392 gggacggatacgaacaccatcccgccgacggctatgacgggccacagggcggcggctatc 451
          |||| ||| ||||||| |||||||||| | ||| ||||| ||  |||||||   || |
Db  76918 GGGAAGGACACGAACATCATCCCGCCGATGCCTACGACGGCCCGAAGGGCGGCAATTACC 76859

Qy    452 ccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaata 511
          ||   || |  |||||  | ||   ||||| |||  || | || |||   ||||  | ||
Db  76858 CCAAACCTACGGGCGCACGAGACGAATACACCTATCACGTCAACGGCACAGCCCGCAGTA 76799

Qy    512 tccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgccgaccgtttccaca 571
          ||   ||||| | |||||||| |||||||| | | |||||  |  |||||  || |  ||
Db  76798 TCAAACTCAATCCGACCGACACCCGCAGCATCCGGCAACGCATATCCGACAATTACAGCA 76739

Qy    572 atgccggcgctatgctgacgcaaggagtaggcgacggattcaaacgcgccacccgataca 631
          |  ||||   |  | |  | | |  || |   || |         |    |||
Db  76738 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 76679

Qy    632 gccccgagctggacagatcgggcaatgccgccgaagccttcaacggcactgcagatatcg 691
            || |||| ||| | | ||||||   |   ||     ||||||||   || |    ||
Db  76678 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 76619

Qy    692 tcaaaaacatcatcggcgcggcaggagaaattgtcggcgcaggcgatgccgtgcagggta 751
            |||  | |   |  || || |||    | |||  ||||||     || | || |
Db  76618 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCTTGGGCATAGGCGACATACTGTACGGAA 76559

Qy    752 taagcgaaggctcaaacattgctgtcatgcacggcttgggtctgctttccaccgaaaaca 811
          || | | |  | |||  || |  |||| |  | | |   |   |  ||  |||   ||
Db  76558 CGCGCTATGCCATAGACAAAGCCGCAATGCGCAACATCGCCCCCTTGCCCGCCGAGGGCA 76499
```

```
Db  76558 CGCGCTATGCCATAGACAAAGCCGCAATGCGCAACATCGCCCCCTTGCCCGCCGAGGGCA 76499

Qy    812 agatggcgcgcatcaacgatttggcagatatggcgcaactcaaagactatgccgcagcag 871
          |  | ||   ||||  ||  ||||      |||||    |  || |  || |    | ||
Db  76498 AATTCGCCGTCATCGGCGGCTTGGGCAGCGTGGCGGGCTTTGAAAAGAATACGCGCGAAG 76439

Qy    872 ccatccgcgattgggcagtccaaaaccccaatgccgcacaaggcatagaagccgtcagca 931
          || |   |   |||  |    |||| ||||||||||| ||  | | |||||||||  ||
Db  76438 CCGTTGACCGGTGGATACAGGAAAATCCCAATGCCGCCGAAACCGTCGAAGCCGTCTTCA 76379

Qy    932 atatctttatggcagcca 949
          |  |       |||||||
Db  76378 ACGTTGCCGCAGCAGCCA 76361


RESULT  27
NMA7Z2491
LOCUS       NMA7Z2491             195767 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 7/7.
ACCESSION   AL162758 AL157959
VERSION     AL162758.2  GI:7380672
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 195767)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 195767)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .195767
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            80. .952
                     /gene="NMA2045"
     CDS             80. .952
                     /gene="NMA2045"
                     /note="NMA2045, possible oxidoreductase, len: 290aa;
                     similar to many eg. SW:P20692 (TYRA_BACSU) prephenate
                     dehydrogenase from Bacillus subtilis (372 aa) fasta
                     scores; E(): 2.8e-30, 39.4% identity in 284 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="putative oxidoreductase"
                     /protein_id="CAB85263.1"
                     /db_xref="GI:7380673"
                     /db_xref="SPTREMBL:Q9JT12"
                     /translation="MPILNHIALIGVGLIGGSFVLDLKRQGLVRTVTGIDTDRDNLER
                     ALERGVIDRASVAIDADSIGGADLVLIATPVATVPAILTALRPVLPEHTWISDVGSTK
                     SSVIEAFRRCLPDRLHRCIAAHPIAGSDRSGAQAAQFGLFRHRKLIITPHGGEHSDGI
                     ALIENLWHAVGAEIYTMDAQRHDAVFAAVSHMPHLTAFAYVHQILDHPDGQEYLKFAA
                     TGFRDFTRIASGHPAVWADICLANKDSLLQLVQGLGKQLDVLADILTADDREALYRYF
                     EEAKTTRDRWLDGN"
     misc_feature    complement(547. .556)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    960. .969
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(1009. .2211)
                     /gene="NMA2046"
     CDS             complement(1009. .2211)
                     /gene="NMA2046"
                     /note="NMA2046, hypothetical integral membrane protein,
                     len: 400aa; similar to TR:P95556 (EMBL:Z73914)
                     hypothetical protein from Pseudomonas stutzeri (396 aa)
                     fasta scores; E(): 0, 35.0% identity in 391 aa overlap.
                     Contains possible membrane-spanning hydrophobic regions."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical integral membrane protein"
                     /protein_id="CAB85264.1"
                     /db_xref="GI:7380674"
                     /db_xref="SPTREMBL:Q9JT11"
                     /translation="MIPFCGKYINDKQGTHMKFTKHPVWAMAFRPFYSLAALYGALSV
                     LLWGFGYTGTHELSGFYWHAHEMIWGYAGLVVIAFLLTAVATWTGQPPTRGGVLVGLT
                     IFWLAARIAAFIPGWGASASGILGTLFFWYGAVCMALPVIRSQNQRNYVAVFALFVLG
                     GTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLIGTRIISFFTSKRLNVPQIPSPK
                     WVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFTVQVYRWWYKPVLKEPMLWIL
                     FAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLTLGMMARTALGHTGNPIYP
                     PPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFALALLVYAWKYIPWLIR
                     PRSDGRPG"
     gene            complement(2208. .2576)
                     /gene="NMA2047"
     CDS             complement(2208. .2576)
                     /gene="NMA2047"
                     /note="NMA2047, hypothetical protein, len: 122aa; weakly
                     similar to SW:P71932 (YQ03_MYCTU) hypothetical protein
                     from Mycobacterium tuberculosis (161 aa) fasta scores;
                     E(): 0.017, 32.5% identity in 117 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB85265.1"
                     /db_xref="GI:7380675"
                     /db_xref="SPTREMBL:Q9JT10"
                     /translation="MKPLKRHPALIGLSRDHHHSLSLCVRLLRTPEERHRDELEPHFS
                     ELETHFREEEAKFAPIWQNVAPELKQRFEKDHARLRQMMASPEYGNAAWNTAFATTLR
                     DHARFEERELFPAVEPFLPA"
     misc_feature    complement(2622. .2631)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            2750. .3193
                     /gene="NMA2048"
     CDS             2750. .3193
                     /gene="NMA2048"
                     /note="NMA2048, conserved hypothetical protein, len:
                     147aa; similar to many eg. SW:P21498 (YJEB_ECOLI)
                     hypothetical protein from Escherichia coli (141 aa) fasta
                     scores; E(): 5.9e-23, 47.9% identity in 140 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAB85266.1"
                     /db_xref="GI:7380676"
                     /db_xref="SPTREMBL:Q9JT09"
                     /translation="MYLTQHTDYGLRVLIYTAINDDALVNISTIAVTYGISKSHLMKV
                     VTALVKGGFLHSVRGKGGGLRLASPPERINIGAVVRHLEPMQLVECMGENNECLITPS
                     CRLTGILGGAMKSFFTYLDGFTLQDLLNKPTYDLLYEPRIPIAVQ"
     stem_loop       3275. .3314
                     /note="Stem loop containing DNA uptake sequences: gcaat
                     gccgtctgaa cagccgatgc ttcagacggc atcgg"
     misc_feature    3280. .3289
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
```

```
misc_feature    complement(3300. .3309)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(3320. .3940)
                /gene="NMA2049"
CDS             complement(3320. .3940)
                /gene="NMA2049"
                /note="NMA2049, conserved hypothetical protein, len:
                206aa; similar to many eg. SW:P45847 (YCIO_ECOLI)
                hypothetical protein from Escherichia coli (206 aa) fasta
                scores; E(): 0, 62.9% identity in 205 aa overlap. Contains
                Pfam match to entry PF01300 Sua5_yciO_yrdC, SUA5/yciO/yrdC
                family and Prosite match to PS01147 SUA5/yciO/yrdC family
                signature."
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAB85267.1"
                /db_xref="GI:7380677"
                /db_xref="SPTREMBL:Q9JT08"
                /translation="MAQFFAIHPDNPQERLIKQAVEIVNKGGVVVYPTDSCYALGCKL
                GDKAAMERILSIRKIDLKHHLTLMCADLSELGTYAKIDNVQFRQLKAATPGPYTFILQ
                ATKDVPARTLHPKRKTIGLRIPDNAIAQALLGELGEPLLSCTLMLPEDGEPLTDPYEI
                RECLEHAVDLVIDGGWCGTEPTTVVDMTDGTELVRQGCGDTAVFGL"
misc_feature    complement(3350. .3877)
                /gene="NMA2049"
                /note="Pfam match to entry PF01300 Sua5_yciO_yrdC,
                SUA5/yciO/yrdC family, score 192.00, E-value 9.6e-54"
misc_feature    complement(3818. .3856)
                /gene="NMA2049"
                /note="PS01147 SUA5/yciO/yrdC family signature."
repeat_unit     4027. .4046
                /note=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
                /label=dRS3
repeat_unit     4047. .4102
                /note="85% identity to consensus
                CCAGTCCGTTCAGTTTCAGTCATTTCCGATAAATTCCTGTTGCTTTTCATTTCTAG"
                /label=RS17
misc_feature    complement(4262. .4271)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(4298. .5497)
                /gene="ackA2"
CDS             complement(4298. .5497)
                /gene="ackA2"
                /EC_number="2.7.2.1"
                /note="NMA2050, ackA2, acetate kinase, len: 399aa; similar
                to many eg. SW:P15046 (ACKA_ECOLI) ackA, acetate kinase
                from Escherichia coli (400 aa) fasta scores; E(): 0, 49.5%
                identity in 400 aa overlap. Also similar to NMA1718 from
                Neisseria meningitidis fasta scores; E(): 0, 63.3%
                identity in 398 aa overlap. Contains Pfam match to entry
                PF00871 Acetate_kinase, Acetokinase family; Prosite match
                to PS01075 Acetate and butyrate kinases family signature 1
                and Prosite match to PS01076 Acetate and butyrate kinases
                family signature 2."
                /codon_start=1
                /transl_table=11
                /product="acetate kinase"
                /protein_id="CAB85268.1"
                /db_xref="GI:7380678"
                /db_xref="SPTREMBL:Q9JT07"
                /translation="MSDQLILVLNCGSSSLKGAVIDRKSGSVVLSCLGERLTTPEAVI
                TFSKDGNKRQVPLSGRNCHAGAVGMLLNELEKHELHDRIQAVGHRIAHGGEKYSESVL
                IDQAVMDELNACIPLAPLHNPANISGILAAQEHFPGLPNVGVMDTSFHQTMPERAYTY
                AVPRELRKKYAFRRYGFHGTSMRYVAPEAACILGKPLEDIRMIIAHLGNGASITAIKN
                GKSVDTSMGFTPIEGLVMGTRCGDIDPGVYSYLTSHAGLDVAQVDEMLNKKSGLLGIS
                ELSNDCRTLEIAADEGHEGARLALEVMTYRLAKYIASMAVGCGGVDALVFTGGIGENS
                RNIRAKTVSYLDFLGLHIDTKANMEKRYGNSGIISPTDSSPAVLVVPTNEELMIACDT
                AELVGIL"
misc_feature    complement(4328. .5485)
```

Query Match 14.3%; Score 269.6; DB 1; Length 195767;

Best Local Similarity 56.3%; Pred. No. 6.5e-53;
Matches 525; Conservative 0; Mismatches 405; Indels 3; Gaps 1;

```
Qy      35 ccatactggcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgatccct 94
           || ||  ||| |     ||      ||  || | |||  | || |||||  | || || |
Db   79040 CCGTAGCGGCGGCCGCACTCATACAGCCCGCCCTCGCGGCGGACTTGGCGCAAGACCCGT 79099

Qy      95 ttatccggcaggttctcgaccgtcagcatttcgaacccgacgggaaataccacctattcg 154
           | ||     |       | | || ||||| | ||||||||  || ||||||||||| ||||
Db   79100 TCATTACCGATAACGCCCAACGGCAGCACTACGAACCCGGAGGCAAATACCACCTCTTCG 79159

Qy     155 gcagcaggggggagcttg---ccnagcgcaacggccatatcggattgggaaacatacaaa 211
           ||  |  | | |     |   || | |||| ||| || |||    |     |     | |
Db   79160 GCGACCCGCGCGGCAGCGTCTCCGACCGCACCGGTCAAATCAACGTCATCCAAGACTATA 79219

Qy     212 gccatcagttgggccacctgatgattcaacaggcggccgttgaaggaaatatcggctaca 271
           |||  | | ||||| ||||| | || || |||||   | |  | || |  |||||||||
Db   79220 CCCACCGGATGGGCAACCTGCTCATCCAGCAGGCAAACATCAACGGCACAATCGGCTACC 79279

Qy     272 ttgtccgcttttccgatcacgggcacaaattccattcgcccttcgacaaccatgcctcac 331
              ||||||||||  |||||  || ||   ||  | ||||||||||||||| ||| |
Db   79280 ACACCCGCTTTTCCGGACACGGATACGAAGAACACGCCCCCTTCGACAACCACGCCGCCG 79339

Qy     332 attccgattctgacgaagccggtagtcccgttgacggattcagcctttaccgcatccatt 391
           |  ||     || |||   || |      |  || || || | | | |||||  || | |
Db   79340 ACAGCGCGAGCGAAGAAAAAGGCAACGTTGACGAAGGCTTTACCGTATACCGGCTCAACT 79399

Qy     392 gggacggatacgaacaccatcccgccgacggctatgacgggccacagggcggcggctatc 451
           |||| ||| |||||||| |||||||||| | ||| |||||  ||  |||||||||   || |
Db   79400 GGGAAGGACACGAACATCATCCCGCCGATGCCTACGACGGCCCGAAGGGCGGCAATTACC 79459

Qy     452 ccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaata 511
           ||   || |  || ||  | ||   ||||| |||  || | || |||   || |  |  |
Db   79460 CCAAACCTACGGGTGCACGCGACGAATACACCTATCACGTCAACGGCACAGCACGCAGCA 79519

Qy     512 tccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgccgaccgtttccaca 571
           ||   ||||| | |||||||| |||||||| | | |||||  |  |||||  || |  ||
Db   79520 TCAAACTCAATCCGACCGACACCCGCAGCATCCGGCAACGCATATCCGACAATTACAGCA 79579

Qy     572 atgccggcgctatgctgacgcaaggagtaggcgacggattcaaacgcgccacccgataca 631
           |  ||||   |  |  |  |  | |  |  || |    || |        |   |||
Db   79580 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 79639

Qy     632 gccccgagctggacagatcgggcaatgccgccgaagccttcaacggcactgcagatatcg 691
           || |||| ||| | | ||||||   |   ||     |||||||   || |    ||
Db   79640 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 79699

Qy     692 tcaaaaacatcatcggcgcggcaggagaaattgtcggcgcaggcgatgccgtgcagggta 751
             |||  | |   |  || || |||    | |||  ||||||     || | || |
Db   79700 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCTTGGGCATAGGCGACATACTGTACGGAA 79759

Qy     752 taagcgaaggctcaaacattgctgtcatgcacggcttgggtctgctttccaccgaaaaca 811
           || | | |  | |||  || |  |||| |  | | |  |   |  || |||| ||
Db   79760 CGCGCTATGCCATAGACAAAGCCGCAATGCGCAACATCGCCCCCTTGCCCGCCGAGGGCA 79819

Qy     812 agatggcgcgcatcaacgatttggcagatatggcgcaactcaaagactatgccgcagcag 871
           |  | ||   ||||  ||  ||||      |||||    |  || |  || |    | ||
Db   79820 AATTCGCCGTCATCGGCGGCTTGGGCAGCGTGGCGGGCTTTGAAAAAAATACGCGCGAAG 79879

Qy     872 ccatccgcgattgggcagtccaaaaccccaatgccgcacaaggcatagaagccgtcagca 931
           || |   |  |||  |    ||||||||||||||||  ||  | | ||||||  |   ||
Db   79880 CCGTTGACCGGTGGATACAGGAAAACCCCAATGCCGCCGAAACCGTCGAAGCCCTGGTCA 79939

Qy     932 atatctttatggcagccatccccatcaaaggga 964
           |  || |   |  ||||   ||  ||   ||
Db   79940 ACGTCCTGCCGTTTGCCAAAGTCAAAAACCTGA 79972
```

```
RESULT  28
AE002420
LOCUS       AE002420               10057 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 62 of 206 of
            the complete genome.
```

```
ACCESSION   AE002420 AE002098
VERSION     AE002420.1  GI:7225876
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10057)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10057)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10057
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            126. .350
                     /gene="NMB0649"
     CDS             126. .350
                     /gene="NMB0649"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41070.1"
                     /db_xref="GI:7225877"
                     /translation="MRLATKDLAEAIRKGQVRKSSFNTEQLRAIEKGESKIPDYTWHH
                     HQDTGRMQLIREGLHHDTGHIGWEAMNKGR"
     gene            323. .742
                     /gene="NMB0650"
     CDS             323. .742
                     /gene="NMB0650"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41071.1"
                     /db_xref="GI:7225878"
                     /translation="MGSNEQRKVTMWKIIKEDSDDLEFAIKCLFSQSIDLNEFKLWIE
                     QVIRDMPIEDIPFYIFDLADFDGGIADIDNIVGFVSSCRLSKSKKNALTGIAFLRGID
                     VYDPPISKEKALKALEKHPEIYQKFQHFFPFVELPPL"
     gene            843. .1232
                     /gene="NMB0651"
     CDS             843. .1232
                     /gene="NMB0651"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41072.1"
                     /db_xref="GI:7225879"
                     /translation="MWKIIKEDSDDLGFAIKCLFSQSIDLNEFKLWIEQVIRDMPIED
                     IPFYIFDLADFDGGIADIDNIVGFVSSCRLSKSKKNALTGIAFLRGIDVYDPPISKEK
                     ALKALEKHPEIYQKFQHFFPFVELPPL"
     gene            1355. .2296
                     /gene="NMB0652"
     CDS             1355. .2296
                     /gene="NMB0652"
                     /note="similar to GP:4838554 percent identity: 79.35;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="mafA protein"
                     /protein_id="AAF41073.1"
                     /db_xref="GI:7225880"
                     /translation="MKTLLLLIPLVLTACGTLTGIPAHGGGKRFAVEQELVAASSRAA
                     VKEMDLSALKGRKAALYVSVMGDQGSGNISGGRYSIDALIRGGYHNNPESATQYSYPA
                     YDTTATTKSDALSSVTTSTSLLNAPAAALTKNSGRKGERSAGLSVNGTGDYRNETLLA
                     NPRDVSFLTNLIQTVFYLRGIEVVPPEYADTDVFVTVDVFGTVRSRTELHLYNAETLK
                     AQTKLEYFAVDRDSRKLLITPKTAAYESQYQEQYALWTGPYKVSKTVKASDRLMVDFS
                     DITPYGDTTAQNRPDFKQNNGKKPDVGNEVIRRRKGG"
     gene            2300. .3568
                     /gene="NMB0653"
     CDS             2300. .3568
                     /gene="NMB0653"
                     /note="MafB-related protein; identified by Oxford;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="MafB-related protein"
                     /protein_id="AAF62314.1"
                     /db_xref="GI:7413436"
                     /translation="MKPLRRLTNLLAACAVAAAALIQPALAADLAQDPFITDNAQRQH
                     YEPGGKYHLFGDPRGSVSDRTGKINVIQDYTHQMGNLLIQQANINGTIGYHTRFSGHG
                     HEEHAPFDNHAADSASEEKGNVDEGFTVYRLNWEGHEHHPADAYDGPKGGNYPKPTGA
                     RDEYTYHVNGTARSIKLNPTDTRSIRQRISDNYSNLGSNFSDRADEANRKMFEHNAKL
                     DRWGNSMEFINGVAAGALNPFISAGEAVDQWMQENPNAAETVEALVNVLPFAKVKNLT
                     KAAKPGKAAVSGDFSDSYKHNTASRLSQSVDGEMFQTRNVDFKAKSIGTKIHDGAQGK
                     HISGHRNYIEGKSTLNQNINPQELLNGIHSGAYPVISKGARGNPVVDFGYPIGSDGKS
                     GLSTNFGTIHSGKNGVHIVPANPKTIKKVQ"
     gene            3571. .3888
                     /gene="NMB0654"
     CDS             3571. .3888
                     /gene="NMB0654"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41074.1"
                     /db_xref="GI:7225881"
                     /translation="MNILPSWLRVGMNIAMLGMIHSDIRLITVDYEEGRRFLKIKNYL
                     SREAITEDHEDMEYLITELWSMCGEYFDEADFECIYSNHSSMELNQINGAVFRRKELI
                     SQA"
     gene            4023. .4730
                     /gene="NMB0655"
     CDS             4023. .4730
                     /gene="NMB0655"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF41075.1"
                     /db_xref="GI:7225882"
                     /translation="MVKTADGYKAIARIRTGDRVFAKDEASGKTGYKPVTARYGNPYQ
                     ETVYIEISDGIGNNQTLISNKIHPFYSQGKWIQAGRLKKGDTLLSESGAKQTVQNITF
                     KQQPLKAYNLTVADWHTYFVKGSQAETEGVWVHNDCPYDKGNQRYKDASYHGKNDNSV
```

```
                KSRAPTNGQAALDNSVQVKSTSPRRVGVDKANNEIVVLNKTQTFNNGSAEYHGHVRSW
                QDLHTDQKNALKKAGLD"
     gene       4756. .5085
                /gene="NMB0656"
     CDS        4756. .5085
                /gene="NMB0656"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF41076.1"
                /db_xref="GI:7225883"
                /translation="MTDKSKTEKLISSDDKQSVIDGILDMVFNSKAYEVPWISEKLME
                LSKNKDLDIAGLSLTCFGHLARLHSNIGDYDKVIPLLHSKQDDPELQGRAEDALEDIS
                LFLSENH"
     gene       5180. .5383
                /gene="NMB0657"
     CDS        5180. .5383
                /gene="NMB0657"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF41077.1"
                /db_xref="GI:7225884"
                /translation="MRTEFFKNPTLAPSTAAPWSKRQTAEKQTPSVVLPFADEYGSTP
                MPPKPSKPPSTLPPPKPQSWQKR"
     gene       5211. .5471
                /gene="NMB0658"
     CDS        5211. .5471
                /gene="NMB0658"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF62315.1"
                /db_xref="GI:7413437"
                /translation="MLLPRQHLGQNGRRLKSKHRPSCCRLRMSTGQPQCRRNRRSRLQ
                HCRRQSRKVGKNGKTGEIKADGRKVNVRIDSTEADLLYPAGQ"
     gene       5502. .5903
                /gene="NMB0659"
     CDS        5502. .5903
                /gene="NMB0659"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"

  Query Match             12.0%;  Score 226.6;  DB 1;  Length 10057;
  Best Local Similarity   58.7%;  Pred. No. 6.8e-43;
  Matches  410;  Conservative    0;  Mismatches  285;  Indels    3;  Gaps    1;

Qy      35 ccatactggcagtgtgcctgccgatgcatgcacacgcctcagatttggcaaacgatccct 94
           || ||  ||| |     ||      ||  || | |||  | || |||||  | || || |
Db    2343 CCGTAGCGGCGGCCGCACTCATACAGCCCGCCCTCGCGGCGGACTTGGCGCAAGACCCGT 2402

Qy      95 ttatccggcaggttctcgaccgtcagcatttcgaacccgacgggaaatacccacctattcg 154
           | ||     |      | | || ||||| | |||||||| ||| ||||||||||| ||||
Db    2403 TCATTACCGATAACGCCCAACGGCAGCACTACGAACCCGGCGGCAAATACCACGTCTTCG 2462

Qy     155 gcagcagggggggagc---ttgccnagcgcaacggccatatcggattgggaaacatacaaa 211
           ||  |  | | |      || || | |||| |||| | |||    |  `  |     | |
Db    2463 GCGACCCGCGCGGCAGCGTTTCCGACCGCACCGGCAAAATCAACGTCATCCAAGACTATA 2522

Qy     212 gccatcagttgggccacctgatgattcaacaggcggccgttgaaggaaatatcggctaca 271
           |||| ||| ||||| ||||| | || ||||||||   | |  | || |  |||||||||
Db    2523 CCCACCAGATGGGCAACCTGCTCATCCAACAGGCAAACATCAACGGCACAATCGGCTACC 2582

Qy     272 ttgtccgcttttccgatcacgggcacaaattccattcgcccttcgacaaccatgcctcac 331
           |||||||||||  ||||| ||| ||   ||  | ||||||||||||| ||| |
Db    2583 ACACCCGCTTTTCCGGACACGGACACGAAGAACACGCCCCCTTCGACAACCACGCCGCCG 2642

Qy     332 attccgattctgacgaagccggtagtcccgttgacggattcagcctttaccgcatccatt 391
           |   ||      || |||   || |     |  || || || | | | |||||  || | |
Db    2643 ACAGCGCGAGCGAAGAAAAAGGCAACGTTGACGAAGGCTTTACCGTATACCGGCTCAACT 2702

Qy     392 gggacggatacgaacaccatcccgccgacggctatgacgggccacagggcggcggctatc 451
           |||| ||| ||||||| |||||||||| | ||| ||||| ||  |||||||||   || |
Db    2703 GGGAAGGACACGAACATCATCCCGCCGATGCCTACGACGGCCCGAAGGGCGGCAATTACC 2762

Qy     452 ccgctcccaaaggcgcgagggatatatacagctacgacataaaaggcgttgcccaaaata 511
           ||   || |  ||||  | ||   |||| ||| || | || |||   ||||  | ||
Db    2763 CCAAACCTACGGGCGCACGAGACGAATACACCTATCACGTCAACGGCACAGCCCGCAGTA 2822

Qy     512 tccgcctcaacctgaccgacaaccgcagcaccggacaacggcttgccgaccgtttccaca 571
           ||   ||||| | |||||||| |||||||| | | |||||  |  |||||  || |  ||
Db    2823 TCAAACTCAATCCGACCGACACCCGCAGCATCCGGCAACGCATATCCGACAATTACAGCA 2882

Qy     572 atgccggcgctatgctgacgcaaggagtaggcgacggattcaaacgcgccacccgataca 631
           |  ||||  |  |  | |  | | |  | || |   || |       |   |||
Db    2883 ACCTCGGCAGCAATTTCTCCGACCGCGCCGATGAAGCCAACAGAAAAATGTTCGAGCACA 2942

Qy     632 gccccgagctggacagatcgggcaatgccgccgaagccttcaacggcactgcagatatcg 691
           || |||| ||| | | ||||||   |   ||    ||||||||   || |    ||
Db    2943 ATGCCAAGCTCGACCGCTGGGGCAACAGCATGGAGTTTATCAACGGCGTCGCCGCCGGCG 3002

Qy     692 tcaaaaacatcatcggcgcggcaggagaaattgtcggc 729
           |||  | |   |   || || |||   || | |
Db    3003 CGCTCAACCCCTTTATCAGCGCGGGCGAAGCCGTTGAC 3040


RESULT  29
A96138
LOCUS       A96138                  1497 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 171 from Patent WO9924578.
ACCESSION   A96138
VERSION     A96138.1  GI:6779961
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1497)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 171 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1497
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      475 a    319 c    365 g    338 t
ORIGIN


  Query Match             6.3%;  Score 119.8;  DB 6;  Length 1497;
  Best Local Similarity   66.4%;  Pred. No. 1.1e-17;
  Matches  172;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy     239 aacaggcggccgttgaaggaaatatcggctacattgtccgctttttccgatcacgggcaca 298
           ||| | | |   |||||||   |||||||||      ||  ||||| |  ||||| |||
Db     245 AACGGACAGGATTTGAAGGTGTTATCGGCTATGAAACCCATTTTTCAGGACACGGACACG 304

Qy     299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
           || | ||    || |||||| || ||||  ||| |    |  ||||||      |||  |
Db     305 AAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGCG 364

Qy     359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
             |  ||| || | | ||||||   | ||| ||   || |  |||   ||||||||  |
Db     365 TAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGCAG 424

Qy     419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
```

```
          |||| |||||||||||| || |||||||| ||||| |  ||  |||| || ||||||||||
Db    425 ACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATATAT 484

Qy    479 acagctacgacataaaagg 497
          |||||||| | || |||||
Db    485 ACAGCTACCATATCAAAGG 503


RESULT  30
A61829
LOCUS       A61829                 1744 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 9 from Patent WO9711181.
ACCESSION   A61829
VERSION     A61829.1  GI:3715998
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 1744)
  AUTHORS   Meyer,T.F., Rudel,T., Scheuerpflug,I., Fischer, Eckhard and
            Maier,J.
  TITLE     NUCLEIC ACID MOLECULES WHICH CODE PROTEINS WHICH MEDIATE THE
            ADHESION OF NEISSERIA CELLS TO HUMAN CELLS
  JOURNAL   Patent: WO 9711181-A 9 27-MAR-1997;
            MAX PLANCK GESELLSCHAFT (DE)
COMMENT     Other publication DE 19534579 970320.
FEATURES             Location/Qualifiers
     source          1. .1744
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
BASE COUNT      560 a    383 c    405 g    396 t
ORIGIN


  Query Match             6.3%;  Score 119.8;  DB 6;  Length 1744;
  Best Local Similarity   66.4%;  Pred. No. 1.2e-17;
  Matches  172;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
          ||| | | |   ||||||| | |||||| ||      ||  ||||| |  ||||| |||
Db    286 AACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACACG 345

Qy    299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
          || | ||    || ||||| || |||| ||| |   |  ||||||      |||  |
Db    346 AAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGCG 405

Qy    359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
            | | ||| || | | ||||||   | ||| ||   || |  |||   ||||||||
Db    406 TAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGCAG 465

Qy    419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
          |||| |||||||||||| || |||||||| ||||| |  ||  |||| || ||||||||||
Db    466 ACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATATAT 525

Qy    479 acagctacgacataaaagg 497
          |||||||| | || |||||
Db    526 ACAGCTACCATATCAAAGG 544


RESULT  31
A61821
LOCUS       A61821                 3287 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 1 from Patent WO9711181.
ACCESSION   A61821
VERSION     A61821.1  GI:3715995
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 3287)
  AUTHORS   Meyer,T.F., Rudel,T., Scheuerpflug,I., Fischer, Eckhard and
            Maier,J.
  TITLE     NUCLEIC ACID MOLECULES WHICH CODE PROTEINS WHICH MEDIATE THE
            ADHESION OF NEISSERIA CELLS TO HUMAN CELLS
  JOURNAL   Patent: WO 9711181-A 1 27-MAR-1997;
            MAX PLANCK GESELLSCHAFT (DE)
COMMENT     Other publication DE 19534579 970320.
FEATURES             Location/Qualifiers
     source          1. .3287
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
                     /clone_lib="GENOMISCHE BIBLIOTHEK IN PBA"
                     /clone_lib="(B) CLON(E): H1967/PES25"
BASE COUNT     1017 a    740 c    752 g    778 t
ORIGIN


  Query Match             6.3%;  Score 119.8;  DB 6;  Length 3287;
  Best Local Similarity   66.4%;  Pred. No. 1.3e-17;
  Matches  172;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
          ||| | | |   ||||||| | |||||| ||      ||  ||||| |  ||||| |||
Db   1829 AACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACACG 1888

Qy    299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
          || | ||    || ||||| || |||| ||| |   |  ||||||      |||  |
Db   1889 AAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGCG 1948

Qy    359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
            | | ||| || | | ||||||   | ||| ||   || |  |||   ||||||||
Db   1949 TAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGCAG 2008

Qy    419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
          |||| |||||||||||| || |||||||| ||||| |  ||  |||| || ||||||||||
Db   2009 ACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATATAT 2068

Qy    479 acagctacgacataaaagg 497
          |||||||| | || |||||
Db   2069 ACAGCTACCATATCAAAGG 2087


RESULT  32
AF142582
LOCUS       AF142582               3294 bp    DNA     linear   BCT 18-MAY-1999
DEFINITION  Neisseria gonorrhoeae maf gene cluster, complete sequence.
ACCESSION   AF142582
VERSION     AF142582.1  GI:4838552
KEYWORDS    .
SOURCE      Neisseria gonorrhoeae.
  ORGANISM  Neisseria gonorrhoeae
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 3294)
  AUTHORS   Eickernjaeger,S., Meyer,T.F., Fischer,E., Maier,J., Manning,P.A.,
            Rudel,T., Scheuerpflug,I., Schulz,E. and Schwan,E.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-1999) Molekulare Biologie, MPI fuer
            Infektionsbiologie, Monbijoustr. 2, Berlin 10117, Germany
FEATURES             Location/Qualifiers
     source          1. .3294
                     /organism="Neisseria gonorrhoeae"
                     /strain="MS11"
                     /db_xref="taxon:485"
     gene            143. .457
                     /gene="mafI"
     CDS             143. .457
                     /gene="mafI"
                     /codon_start=1
                     /transl_table=11
                     /product="MafI"
                     /protein_id="AAD31037.1"
```

```
                     /db_xref="GI:4838553"
                     /translation="MRKSRLSRYKQNKLIELFVAGVTARTAAEPDSIVYTDCYRRYDV
                     LDAGEFSHFRINHSTHFAERQNHINGIGNFWNRAKRHLRKFDGIPKEHFEPYLKECER
                     RF"
     gene            590. .1552
                     /gene="mafA"
     CDS             590. .1552
                     /gene="mafA"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesin MafA"
                     /protein_id="AAD31038.1"
                     /db_xref="GI:4838554"
                     /translation="MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAAS
                     ARAAVKDMDLQALHGRKVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDY
                     TYPRYETTAETTSGGLTGLTTSLSTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNE
                     TLTTNPRDTAFLSHLVQTVFFLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNA
                     ETLKAQTKLEYFAVDRTNKKLLIKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLM
                     VDFSDIRPYGNHTGNSAPSVEADNSHEGYGYSDEAVRQHRQGQP"
     gene            1592. .3121
                     /gene="mafB"
     CDS             1592. .3121
                     /gene="mafB"
                     /codon_start=1
                     /transl_table=11
                     /product="adhesin MafB"
                     /protein_id="AAD31039.1"
                     /db_xref="GI:4838555"
                     /translation="MNLPIQKFMMLFAAAISLLQIPISHANGLDARLRDDMQAKHYEP
                     GGKYHLFGNARGSVKNRVCAVQTFDATAVGPILPITHERTGFEGIIGYETHFSGHGHE
                     VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPADGYDGPQGGGYPEPQGARD
                     IYSYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNW
                     RANRMDDIRGIVQGAVNPFLTGFQGLGVGAITDSAVSPVTYAAARKTLQGIHNLGNLS
                     PEAQLAAASLLQDSAFAVKDGINSARQWADAHPNITATAQTALAVAEAAGTVWGGKKV
                     ELNPTKWDWVKNTGYEKPAARPMQTVDGEMAGKNKPPKPSTQQHSTHSDNNIGLPAPY
                     VKPDTSISPTGTIQDRIRWTKSKFPTEKSLNGHFKAHGKEFGDITIEDYQKMASDLLS
                     KQTSDKILGYQTEHRRVRYDINNNIYVLANPKTFKIKTMFKPNLGKEYYDGEFKKDMG
                     N"
BASE COUNT     1017 a    743 c    755 g    779 t
ORIGIN


  Query Match             6.3%;  Score 119.8;  DB 1;  Length 3294;
  Best Local Similarity   66.4%;  Pred. No. 1.3e-17;
  Matches  172;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
          ||| | | |   ||||||| | |||||| ||      ||  ||||| |  ||||| |||
Db   1836 AACGGACAGGGTTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACACGGACACG 1895

Qy    299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
          || | ||    || ||||| || ||||  ||| |   |  ||||||     |||  |
Db   1896 AAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGCG 1955

Qy    359 ccgttgacggattcagccttttaccgcatccattgggacggatacgaacaccatcccgccg 418
            |  | ||| || | | ||||||   | ||| ||   || |  |||   |||||||| |
Db   1956 TAGACGGCGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATACATCCCGCAG 2015

Qy    419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
          |||| ||||||||||| || |||||||| ||||| |  ||  |||| || ||||||||||
Db   2016 ACGGATATGACGGGCCTCAAGGCGGCGGTTATCCGGAACCACAAGGGGCAAGGGATATAT 2075

Qy    479 acagctacgacataaaagg 497
          |||||||| | || |||||
Db   2076 ACAGCTACCATATCAAAGG 2094


RESULT  33
NMA1Z2491/c
LOCUS       NMA1Z2491             340806 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 1/7.
ACCESSION   AL162752 AL157959
VERSION     AL162752.2  GI:7378778
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 340806)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 340806)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .340806
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            complement(38. .358)
                     /gene="NMA0001"
     CDS             complement(38. .358)
                     /gene="NMA0001"
                     /note="NMA0001, unknown, len: 106 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0001"
                     /protein_id="CAB83321.1"
                     /db_xref="GI:7378779"
                     /db_xref="SPTREMBL:Q9JX93"
                     /translation="MSNTQTIRSADFTTSRAWGALDIANMNGTTVRLHWTDQPYKWHV
                     NDGEEVFAVMDGEVDMHYRENGEEHIVRLKSGDIFYAGIGTEHVAHPRGETRILVIEK
                     EGSV"
     gene            complement(414. .2438)
                     /gene="nuoL"
     CDS             complement(414. .2438)
                     /gene="nuoL"
                     /EC_number="1.6.5.3"
                     /note="NMA0002, nuoL, NADH dehydrogenase I chain L, len:
                     674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase
                     I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to
                     entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone
                     (complex I), various chains, and to entry PF00662
                     oxidored_q1_N, NADH-Ubiquinone oxidoreductase (complex I),
                     chain 5 N-terminus"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain L"
                     /protein_id="CAB83322.1"
                     /db_xref="GI:7378780"
                     /db_xref="SWISS-PROT:Q9JX92"
                     /translation="MNDMTLYLIIALVPLAGSLIAGLFGNKIGRAGAHTVTILGVAVS
                     AVLSAYVLWGFLNGSRAKFDENVYTWLTMGGLDFSVGFLVDTMTAMMMVVVTGVSLMV
                     HIYTIGYMHDEKVGYQRFFSYISLFTFSMLMLIMSNNFIQLFFGWEAVGLVSYLLIGF
                     YFKRPSATFANLKAFLINRVGDFGFLLGIGLVLAYFGGSLRYQDVFAYLPNVQNATIQ
                     LFPGVEWSLITVTCLLLFVGAMGKSAQFPLHVWLPDSMEGPTPISALIHAATMVTAGL
                     FMVSRMSPIYEMSSTALSVIMVIGAITALFMGFLGVIQNDIKRVVAYSTLSQLGYMTV
                     ALGASAYSVAMFHVMTHAFFKALLFLAAGSAIIGMHHDQDMRHMGNLKKYMPITWLTM
                     LIGNLSLIGTPFFSGFYSKDSIIEAAKYSTLPGSGFAYFAVLASVFVTAFYAFRQYFM
```

```
                VFHGEEKWRSLPEHHSDGHGEEHHGLGKNDNPHESPLVVTLPLILLAVPSVIIGYIAI
                EPMLYGDFFKDVIFVNADAHPTMHIMKEEFHGALAMVSHSLHSPVLYLAIAGVLSAWL
                LYVKLPHLPAKIAQAFRPVYVLFENKYYLDALYFNVFAKGTRALGTFFWKVGDTAIID
                NGIVNGSARLVGAVAAQVRKVQTGFIYTYAAAMVFGVLVLLGMTFWGLFR"
misc_feature    1035. .1044
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    complement(1116. .2030)
                /gene="nuoL"
                /note="Pfam match to entry PF00361 oxidored_q1,
                NADH-Ubiquinone/plastoquinone (complex I), various chains,
                score 351.80, E-value 7.3e-102"
misc_feature    complement(2061. .2243)
                /gene="nuoL"
                /note="Pfam match to entry PF00662 oxidored_q1_N,
                NADH-Ubiquinone oxidoreductase (complex I), chain 5
                N-terminus, score 77.20, E-value 3.5e-19"
gene            complement(2487. .2828)
                /gene="NMA0003"
CDS             complement(2487. .2828)
                /gene="NMA0003"
                /note="NMA0003, unknown, len: 113 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0003"
                /protein_id="CAB83323.1"
                /db_xref="GI:7378781"
                /db_xref="SPTREMBL:Q9JX91"
                /translation="MMDKNQLEQEFHKAMLNIYQEALNLPQPYKATRFLQIVNEFGGK
                EAADKLLSTGEKKTQTGFTELILSGGGVHALKYSMEYLVLQKPWCDLFTEEQLAVARK
                RLERVGFVSPK"
RBS             complement(2833. .2837)
gene            complement(2952. .3527)
                /gene="NMA0004"
CDS             complement(2952. .3527)
                /gene="NMA0004"
                /note="NMA0004, unknown, len: 191 aa; similar to
                hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa),
                fasta scores; E(): 0; 80.5% identity in 190 aa overlap,
                and to proteins involved in cell cycle e.g. TR:O54679
                (EMBL:AF036487) Lactococcus lactis Plasmid pNZ4000
                putative mobilization protein (200 aa), fasta scores; E():
                0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell
                filamentation protein FIC (200 aa), fasta scores; E():
                0.093, 29.4% identity in 126 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0004"
                /protein_id="CAB83324.1"
                /db_xref="GI:7378782"
                /db_xref="SPTREMBL:Q9JQR9"
                /translation="MPSENPIGKTMKSIDEQSLHNARRLFESGDIDRIEVGTTAGLQQ
                IHRYLFGGLYDFAGQIREDNISKGGFRFANAMYLKEALVKIEQMPERTFEEIIAKYVE
                MNIAHPFLEGNGRSTRIWLDLVLKKNLKKVVNWQNVSKTLYLQAMERSPVNDLELRFL
                LKDNLTDDVDNREIIFKGIEQSYYYEGYEKG"
misc_feature    complement(3516. .3525)
                /gene="NMA0004"
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(3561. .3866)
                /gene="nuoK"
CDS             complement(3561. .3866)
                /gene="nuoK"
                /EC_number="1.6.5.3"
                /note="NMA0005, nuoK, NADH dehydrogenase I chain K, len:
                101 aa; simlar to many e.g. NUOK_RHOCA NADH dehydrogenase
                I chain K (EC 1.6.5.3) (102 aa), fasta scores; E():
                1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam
                match to entry PF00420 oxidored_q2,
                NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
                /codon_start=1
                /transl_table=11
                /product="NADH dehydrogenase I chain K"
                /protein_id="CAB83325.1"
                /db_xref="GI:7378783"
                /db_xref="SPTREMBL:Q9JQU9"
                /translation="MITLTHYLVLGALLFGISAMGIFMNRKNVLVLLMSIELMLLAVN
                FNFIAFSQHLGDTAGQIFVFFVLTVAAAESAIGLAIMVLVYRNRQTINVADLDELKG"
misc_feature    complement(3564. .3863)
                /gene="nuoK"
                /note="Pfam match to entry PF00420 oxidored_q2,
                NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
                score 141.50, E-value 1.5e-38"
gene            complement(3863. .4534)
                /gene="nuoJ"
CDS             complement(3863. .4534)
                /gene="nuoJ"
                /EC_number="1.6.5.3"
                /note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
                223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
                I chain J (EC 1.6.5.3) (202 aa), fatsa scores; E():
                9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
                match to entry PF00499 oxidored_q3,
                NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
                /codon_start=1
                /transl_table=11
                /product="NADH dehydrogenase I chain J"
                /protein_id="CAB83326.1"
                /db_xref="GI:7378784"
                /db_xref="SPTREMBL:Q9JX90"
                /translation="MTFQLILFYIFAVIILYGAIKTVTAKNPVHAALHLVLTFCVSAM
                LWMLMQAEFLGVTLVVVYVGAVMVLFLFVVMMLNIDIEEMRAGFWRHAPVAGVVGTLL
                AVALILILVNPKTDLAAFGLMKDIPADYNNIRDLGSRIYTDYLLPFELAAVLLLLGMV
                AAIALVHRKTVNPKRMDPADQVKVRADQGRMRLVKMEAVKPQTESAEESEVSDDLKPK
                EEGKA"
RBS             complement(3874. .3879)
                /gene="nuoJ"
misc_feature    complement(4034. .4525)
                /gene="nuoJ"
                /note="Pfam match to entry PF00499 oxidored_q3,
                NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
                score 77.90, E-value 2e-19"
gene            complement(4570. .5277)
                /gene="NMA0007"
CDS             complement(4570. .5277)
                /gene="NMA0007"
                /note="NMA0007, unknown, len: 235 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0007"
                /protein_id="CAB83327.1"
                /db_xref="GI:7378785"
                /db_xref="SPTREMBL:Q9JX89"
```

```
Query Match             5.7%;  Score 108.2;  DB 1;  Length 340806;
Best Local Similarity   62.3%;  Pred. No. 1.4e-14;
Matches  170;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy     239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
           ||| | | |   ||||||| | |||||| ||      ||  ||||| |  || || ||
Db  305685 AACGGACAGGATTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACATGGACATG 305626

Qy     299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
           || | ||    || ||||| || ||||  ||| |   |  ||||||      |||  |
Db  305625 AAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGCG 305566

Qy     359 ccgttgacggattcagccttaccgcatccattgggacggatacgaacaccatcccgccg 418
             |  |  || || | | ||||||   | ||| ||   || |  |||  |||||| |  |
Db  305565 TAGACGGTGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATCCATCCGGAGG 305506

Qy     419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
           | || |||||||||||| || ||| |||  |||||  | |||  ||| || ||||||||||
Db  305505 ATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATATAT 305446

Qy     479 acagctacgacataaaaggcgttgcccaaaata 511
           |||||||| |  | |||||   | |   ||| |
```

```
Db 305445 ACAGCTACTATGTCAAAGGAACTTCAACAAAAA 305413


RESULT 34
A96134
LOCUS       A96134                  1449 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 167 from Patent WO9924578.
ACCESSION   A96134
VERSION     A96134.1  GI:6779959
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1449)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 167 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1449
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      445 a    310 c    335 g    340 t     19 others
ORIGIN


  Query Match             5.6%;  Score 105.2;  DB 6;  Length 1449;
  Best Local Similarity   61.2%;  Pred. No. 3.3e-14;
  Matches  167;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy    239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
          ||| | | |   |||||||  | ||||||  ||      ||  ||||| |  || ||  ||
Db    245 AACGGACAGGATTTGAAGGCATTATCGGTTATGAAACCCATTTTTCAGGACATGGACATG 304

Qy    299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
          || | ||    || ||||| || ||||  ||| |    |  ||||||       |||  |
Db    305 AAGTACACAGTCCGTTCGATAATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGCG 364

Qy    359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
            |  |  || || | | ||||||   | ||| ||   || |  |||  ||||||  |  |
Db    365 TAGACGGTGGTTTTACCGTTTACCAACTTCATCGGACAGGGTCGGAAATCCATCCGGAGG 424

Qy    419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
          | || |||||||||| || ||| |||  |||||  | |||  ||| || ||||||||||
Db    425 ATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATATAT 484

Qy    479 acagctacgacataaaaggcgttgcccaaaata 511
          |||  ||  |  | |||||   | |   ||| |
Db    485 ACANNTANTATGTCAAAGGAACTTCAACAAAAA 517


RESULT 35
A96132
LOCUS       A96132                  1452 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 165 from Patent WO9924578.
ACCESSION   A96132
VERSION     A96132.1  GI:6779958
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1452)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 165 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1452
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      466 a    298 c    332 g    356 t
ORIGIN


  Query Match             5.6%;  Score 105;  DB 6;  Length 1452;
  Best Local Similarity   61.5%;  Pred. No. 3.6e-14;
  Matches  168;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy    239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
          ||| | | |   |||||||  ||||||  ||      ||  ||||| |  ||||| ||
Db    245 AACGGACAGGGTTTGAAGGTGTTATCGGTTATGAAACCCATTTTTCAGGGCACGGACATG 304

Qy    299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
          || | ||    || |||||  | ||||  ||| |    |  ||||||       |||  ||
Db    305 AAGTACACAGTCCGTTCGATCATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGTG 364

Qy    359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
            |  | ||| || |   ||||||   | ||| |    || |  |||  ||||||  |  |
Db    365 TAGACGGCGGTTTTACTGTTTACCAACTTCATCGAACAGGGTCGGAAATCCATCCGGAGG 424

Qy    419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
          | || |||||||||| || ||| |||  |||||  | |||  ||| || ||||||||||
Db    425 ATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATATAT 484

Qy    479 acagctacgacataaaaggcgttgcccaaaata 511
          |||||||  |  | |||||   | |   ||| |
Db    485 ACAGCTATTATGTCAAAGGAACTTCAACAAAAA 517


RESULT 36
AE002559
LOCUS       AE002559                10869 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 201 of 206
            of the complete genome.
ACCESSION   AE002559 AE002098
VERSION     AE002559.2  GI:7413482
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10869)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10869)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Apr 4, 2000 this sequence version replaced gi:7227357.
FEATURES             Location/Qualifiers
     source          1. .10869
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
```

/db_xref="taxon:122586"
/note="serogroup: B"
gene 150. .524
/gene="NMB2095"
CDS 150. .524
/gene="NMB2095"
/note="similar to PID:1817528 percent identity: 53.99; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="adhesin complex protein, putative"
/protein_id="AAF42412.1"
/db_xref="GI:7227358"
/translation="MKLLTTAILSSAIALSSMAAAAGTDNPTVAKKTVSYVCQQGKKV
KVTYGFNKQGLTTYASAVINGKRVQMPVNLDKSDNVETFYGKEGGYVLGTGVMDGKSY
RKQPIMITAPDNQIVFKDCSPR"
gene complement(712. .2178)
/gene="NMB2096"
CDS complement(712. .2178)
/gene="NMB2096"
/note="similar to SP:P33940 PID:424051 GB:U00096 percent identity: 69.73; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="malate:quinone oxidoreductase"
/protein_id="AAF42413.1"
/db_xref="GI:7227359"
/translation="MAEATDVVLVGGGIMSATLGVLLKELEPSWEITLIERLEDVALE
SSNAWNNAGTGHSALCELNYAPLGANGIIDPARALNIAEQFHVSRQFWATLVAEGKLE
DNSFINAVPHMSLVMNEDHCSYLQKRYDAFKTQKLFENMEFSTDRNKISDWAPLMMRG
RDENQPVAANYSAEGTDVDFGRLTRQMVKYLQGKGVKTEFNRHVEDIKRESDGAWVLK
TADTRNPDGQLTLRTRFLFLGAGGGALTLLQKSGIPEGKGYGGFPVSGLFFRNSNPET
AEQHNAKVYGQASVGAPPMSVPHLDTRNVDGKRHLMFGPYAGFRSNFLKQGSLMDLPL
SIHMDNLYPMLCAGWANMPLTKYLLGELRKTKEERFASLLEYYPEANPDDWELITAGQ
RVQIIKKDSEKGGVLQFGTEIVAHADGSLAALLGASPGASTAVPLMIRLMHQCFPERA
PSWEDRLKELVPGYGIKLNENPERADEIIAYTAKVLDI"
gene complement(2230. .2319)
/gene="NMB2097"
CDS complement(2230. .2319)
/gene="NMB2097"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF42414.1"
/db_xref="GI:7227360"
/translation="MLATHSNPETAYSDTHHCRFSLLLRFRRL"
gene complement(2715. .3305)
/gene="NMB2098"
CDS complement(2715. .3305)
/gene="NMB2098"
/note="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF42415.1"
/db_xref="GI:7227361"
/translation="MSEKIILPVLSLGGVRLEPLDVHHETGLREAVCDGEVWKLGVTS
APHPDRVADYIGTALATRLAFAVVDEEAGRVVGTTAYYHFEPQIPRLDIGFTWYAASA
RRTRINTCCKIMLLDSAFDVLACRCVGWRTDILNLASQRAIERLGAEKDGVLRMHMLR
KDGSVRDTVVYSMLREDWCKNREILTGRLAGYGVQV"
gene complement(3366. .3947)
/gene="NMB2099"
CDS complement(3366. .3947)
/gene="NMB2099"
/note="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF42416.1"
/db_xref="GI:7227362"
/translation="MRNEEKRALRRELRGRRSQMGRDVRAAATVKINHLLKRYIKKGR
KIGVYWPMGKELRLDGFVRAAQKRGAELYLPYIEPRSRRMWFTPYPADGVKQERKRGR
AKLHVPQFAGRKKRVHDLNLLLVPVVGMDRLGYRLGQAGGYYDATLSAMKYRLQAKTV
GVGFACQLVDRLPVEAHDRSLDGFVSEAGILCF"
gene 4025. .4258
/gene="NMB2100"
CDS 4025. .4258
/gene="NMB2100"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF42417.1"
/db_xref="GI:7227363"
/translation="MWHIVAIGYLFVAVMYSAAQPSIARALIYLVFWAVLPTVFTVFT
ITVRRRNHLMRQQEQAESEQQRAQRQKDSGTKP"
gene 4431. .5159
/gene="NMB2101"
CDS 4431. .5159
/gene="NMB2101"
/note="similar to SP:P34831 GB:X53651 PID:296030 percent identity: 78.02; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="30S ribosomal protein S2"
/protein_id="AAF42418.1"
/db_xref="GI:7227364"
/translation="MSQITMRQMIEAGVHFGHQTRFWNPKMAQYIFGARNKIHIVNLE
KTLPMFQDAQEAVRRLVANKGTVLFVGTKRQARDIIREEATRAGMPFVDYRWLGGMLT
NYKTVKQSIKRLEEKTAALENAAESGFSKKEILEMQRDVEKLERSLGGIKNMKGLPDA
IFVIDTGYQKGTLVEAEKLGIPVIAVVDTNNSPDGVKYVIPGNDDSAKAIRLYCRGIA
DAVLEGKNQALQETVAAAQEAAAE"
gene 5288. .6142
/gene="NMB2102"
CDS 5288. .6142
/gene="NMB2102"
/note="similar to GB:D26562 SP:P02997 GB:D13334 GB:V00343 PID:1208943 percent identity: 69.01; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="elongation factor TS (EF-TS)"
/protein_id="AAF42419.1"
/db_xref="GI:7227365"
/translation="MAEITAKMVADLRAATGLGMMECKKALVEAEGNFDKAEEILRIK
SGAKAGKLAGRTAAEGVLAYAINGNVGALVEVNCETDFVAKDAGFVEFANFVAKTAAE
KKPASVEELSELVEAERKAIIAKLGENMSVRRFQVIDTANQLVAYIHGALATEGVLVE
YKGSEDVARKIGMHIVAAKPQCVSEAEVDAETVEKERHIYTEQAIASGKPADIAAKMV
EGRIRKFLAEITLNGQAFVMNPDQTVAQFSKENGTEVISFVRYKVGDGIEKKAVDYAA
EVAAAAKV"
gene 6359. .7078
/gene="NMB2103"
CDS 6359. .7078
/gene="NMB2103"
/note="similar to GB:D26562 SP:P29464 GB:D13334 GB:X78809 PID:1208944 percent identity: 73.16; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="uridylate kinase"
/protein_id="AAF42420.1"
/db_xref="GI:7227366"
/translation="MTQQIKYKRVLLKLSGESLMGSDPFGINHDTIVQTVGEIAEVVK
MGVQVGIVVGGGNIFRGVSAQAGSMDRATADYMGMMATVMNALALKDAFETLGIKARV
QSALSMQQIAETYARPKAIQYLEEGKVVIFAAGTGNPFFTTDTAAALRGAEMNCDVML
KATNVDGVYTADPKKDPSATRYETITFDEALLKNLKVMDATAFALCRERKLNIVVFGI
AKEGSLKRVITGEDEGTLVHC"
gene 7262. .8225
/gene="NMB2104"

```
                     /note="similar to GP:4838554 percent identity: 99.38;
                     identified by sequence similarity; putative"
     gene            8265. .9716
                     /gene="NMB2105"
     CDS             8265. .9716
                     /gene="NMB2105"
                     /note="similar to GP:4838555 percent identity: 0.00;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="mafB protein"

  Query Match             5.6%;  Score 105;  DB 1;  Length 10869;
  Best Local Similarity   61.5%;  Pred. No. 4.8e-14;
  Matches  168;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
           ||| | | |   |||||||   |||||| ||      ||  ||||| |  |||||| ||
Db    8509 AACGGACAGGGTTTGAAGGTGTTATCGGTTATGAAACCCATTTTTCAGGGCACGGACATG 8568

Qy     299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
           || | ||    || |||||  | ||||  ||| |   |  ||||||      |||  ||
Db    8569 AAGTACACAGTCCGTTCGATCATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGTG 8628

Qy     359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
             | | ||| || |   ||||||   | ||| |    || |  |||  |||||| |  |
Db    8629 TAGACGGCGGTTTTACTGTTTACCAACTTCATCGAACAGGGTCGGAAATCCATCCGGAGG 8688

Qy     419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
           | || |||||||||||| || ||| |||  |||||  | |||  ||| || |||||||||| |
Db    8689 ATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATATAT 8748

Qy     479 acagctacgacataaaaggcgttgcccaaaata 511
           |||||||  |  | |||||   | |   ||| |
Db    8749 ACAGCTATTATGTCAAAGGAACTTCAACAAAAA 8781


RESULT  37
AX044035
LOCUS       AX044035               172325 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 114 from Patent WO0066791.
ACCESSION   AX044035
VERSION     AX044035.1  GI:11342919
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 172325)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 114 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .172325
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    43072 a  47583 c  41465 g  40205 t
ORIGIN

  Query Match             5.6%;  Score 105;  DB 6;  Length 172325;
  Best Local Similarity   61.5%;  Pred. No. 7.1e-14;
  Matches  168;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
           ||| | | |   |||||||   |||||| ||      ||  ||||| |  |||||| ||
Db  122939 AACGGACAGGGTTTGAAGGTGTTATCGGTTATGAAACCCATTTTTCAGGGCACGGACATG 122998

Qy     299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
           || | ||    || |||||  | ||||  ||| |   |  ||||||      |||  ||
Db  122999 AAGTACACAGTCCGTTCGATCATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGTG 123058

Qy     359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
             | | ||| || |   ||||||   | ||| |    || |  |||  |||||| |  |
Db  123059 TAGACGGCGGTTTTACTGTTTACCAACTTCATCGAACAGGGTCGGAAATCCATCCGGAGG 123118

Qy     419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
           | || |||||||||||| || ||| |||  |||||  | |||  ||| || |||||||||| |
Db  123119 ATGGATATGACGGGCCGCAAGGCAGCGATTATCCGCCCCCCGGAGGAGCAAGGGATATAT 123178

Qy     479 acagctacgacataaaaggcgttgcccaaaata 511
           |||||||  |  | |||||   | |   ||| |
Db  123179 ACAGCTATTATGTCAAAGGAACTTCAACAAAAA 123211


RESULT  38
A96130
LOCUS       A96130                 375 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 163 from Patent WO9924578.
ACCESSION   A96130
VERSION     A96130.1  GI:6779957
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 163 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      105 a     89 c     90 g     90 t      1 others
ORIGIN

  Query Match             5.1%;  Score 96;  DB 6;  Length 375;
  Best Local Similarity   59.3%;  Pred. No. 4.1e-12;
  Matches  162;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     239 aacaggcggccgttgaaggaaatatcggctacattgtccgcttttccgatcacgggcaca 298
           ||| | | |   |||||||   |||||| ||      ||  ||||| |  |||||| ||
Db      29 AACGGACAGGGTTTGAAGGTGTTATCGGTTATGAAACCCATTTTTCAGGGCACGGACATG 88

Qy     299 aattccattcgcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtc 358
           || | ||    || |||||  | ||||  ||| |   |  ||||||      |||  ||
Db      89 AAGTACACAGTCCGTTCGATCATCATGATTCAAAAAGCACTTCTGATTTCAGCGGCGGTG 148

Qy     359 ccgttgacggattcagcctttaccgcatccattgggacggatacgaacaccatcccgccg 418
             | | ||| || |   ||||||   | ||| |    | |  |||  |||||| |  |
Db     149 TAGACGGCGGTTTTACTGTTTACCAACTTCATCGAACATGGTCGGAAATCCATCCGGAGG 208

Qy     419 acggctatgacgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatat 478
           | |  |||||||||||| || |   |     |||||  | |||  ||| || |||||||||| |
Db     209 ATGAATATGACGGGCCGCAAGCAGCGNATTATCCGCCCCCCGGAGGAGCAAGGGATATAT 268

Qy     479 acagctacgacataaaaggcgttgcccaaaata 511
           |||||||  |  | |||||   | |   ||| |
Db     269 ACAGCTATTATGTCAAAGGAACTTCAACAAAAA 301


RESULT  39
I66494/c
```

```
LOCUS       I66494                  7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

```
  Query Match             4.3%;  Score 80.2;  DB 6;  Length 7218;
  Best Local Similarity   5.5%;  Pred. No. 3.5e-08;
  Matches   22;  Conservative  239;  Mismatches  142;  Indels    0;  Gaps    0;

Qy   1357 gatgcgaaaccgagatgggaggttgataggaagcttaataaattgacaactcgtgagcag 1416
          :: : ::::  :::: ::::::  :: ::::::   :: :::  :: ::   : ::: ::
Db   1424 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1365

Qy   1417 gtggagaaaaatgttcaggaaacgagaagaaggagtcagagtagtcagtttaaagcccat 1476
          : ::::::::: :   :::::: ::::::::::::  :::: ::  ::   :::: :
Db   1364 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1305

Qy   1477 gcgcaacgagaatgggaaaataaaacagggttagattttaatcattttataggtggtgat 1536
          : : :: ::::: :::::::: :::: :::: :::    ::  :    : ::: :: ::
Db   1304 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1245

Qy   1537 atcaataagaaaggcacagtaacaggagggcatagtctaacccgtggtgatgtacgggtg 1596
          :  :: :::::::: : :: :: :::::: : ::   ::   : :: :: : : ::: :
Db   1244 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1185

Qy   1597 atacaacaaacctcggcacctgataaacatggggtttatcaagcgacagtggaaattaaa 1656
          : : :: :::    :: :   :: ::: : ::::   :  ::: :: :: :::::  :::
Db   1184 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1125

Qy   1657 aagcctgatggaagttgggaggtgaaaacgaaaaaaggtgggaaagtgatgaccaagcac 1716
          :::   :: :::::  :::::: ::::: ::::::::: :::::::: :: :: ::: |
Db   1124 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRAT 1065

Qy   1717 accatgttcccaaaagattgggatgaggctagaattagggctg 1759
             || | ||||   |  |   |   ||    ||||||   |||
Db   1064 CGCAAGCTCCCTCGACCTGCAGCCAAGCTCGGAATTAATTCTG 1022
```

```
RESULT  40
AF429315
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT     29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN
```

```
  Query Match             2.6%;  Score 49;  DB 9;  Length 125020;
  Best Local Similarity   10.5%;  Pred. No. 1.3;
  Matches   96;  Conservative  395;  Mismatches  419;  Indels    8;  Gaps    3;

Qy    314 tcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacggattca 373
          :: :: ::|  :|: ::|   | :   : : | : :::   :::|| : ::: :| :|
Db  16975 KMMRSSMMCTSSCYASMCCMCCWSCCCMRSCCCMCRSYCCCMRYCCACKCYMSSYWTMSA 17034

Qy    374 gcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatgacgggc 433
          :::    : :   |:    ::::: :  :: |  ::    |  :  :: ::::: : :
Db  17035 SYMSRRSYWKRSMKCWSRMCSRSRSSKCKSRGCGGSMGKGKGGKSYGRKTKRKSRGKMKG 17094

Qy    434 cacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacgacataa 493
          :::::::::::: |::|:: :  | :    : :|    :| ::   :
Db  17095 AKWYMYYRRSMRKMKMYSSKGMYCMYCWCWGRRGCYCSCMTSRSAMCCSYCYAKCKSMCY 17154

Qy    494 aaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggacaacggc 553
            : :: |  :: :::: :|:: :|   :  |         | ||       |
Db  17155 SCYYGMSMKGYYYACSYRGSMSSKYCMRGSTYSTSCGCCCTTTTTCCCCCNANTGGGGAA 17214

Qy    554 ttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacggattca 613
            | ::|:::   : ||       :    :::|:||   |  ::::  ::  :
Db  17215 GCTTTTNCNKKTYSYYRKRNGCAMCKYNNYYNNSWRSSCRAGMSCTKYKSSMTMSMASYC 17274

Qy    614 aacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagccttca 673
          : ::: :| : : | : :: :::   ::::: :: :::    :: |: ::: :::   :
Db  17275 WCMSMYYCSMRSMASRGMSWSSYMKMMKSSWRMSCYMKCCWCMKMCYCMMRSMRSRGMSY 17334

Qy    674 acggcactgcagatatcgtcaaaaacatcatcggcgcggcaggagaaattgtcggcgcag 733
           :  ::  :::| : |  :  :   : :  ::  ::::::  |    : : :  |: :  :
Db  17335 YMYASWKSSSRGCYCTRCYWCMSSKSCYKSYYMMMRSKRMKGMKWSRCWGSMWGSASRSS 17394

Qy    734 gcgatgccgtgcagggtataagcgaaggctcaaacattgctgtcatg--cacggcttggg 791
          :: ::::: ::::::    ||| :::| ::: :  :    ::  ::  ::::::::|:|
Db  17395 CYKCYKSMRCSMMSSKCYRCAGCMMMKGGYMRYMRCWSMKRRWWGKSAMYRMRRMWKGRG 17454

Qy    792 tctgctttccaccgaaaacaagatggcgcgcatcaacgatttggcagatatggcgcaact 851
```

```
Db  17455 AMMMCMKCYSRMSRCMMMMKSYCASCRSCAMMSGTYMKCASYMCMSSYWCTCWCTSYMSY 17514

Qy    852 caaagactatgccgcagcagccatccgcgattggg----cagtccaaaaccccaatgccg 907
Db  17515 YRCTCWCKGWSSYTKSKSWSSSMSSYKKGRRKSYSMCCTSRGAMSCWRRCCYMRGASSMR 17574

Qy    908 cacaaggcatagaagccgtcagcaatatctttatggcagccatccccatcaaagggattg 967
Db  17575 AGSMSRRRAKGRSWGGRSKWWMTGGMWRSKYYYYCTGRRMMMTYMCC--CWRRRSYMYYR 17632

Qy    968 gagctgtccggggaaaatacggcttgggcggcatcacggcacatcctgtcaagcggtcgc 1027
Db  17633 SMSAMGMRKSSWSGMRMMGSASSRRCKSASRSSWCSRRMKGMRGSCWSSKMWGGSRSRSA 17692

Qy   1028 agatgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaattttgccgatg 1087
Db  17693 SSKCKGSRGMRRRSKRSSKYRKRGRGKKRSMTKSKGSKGKSKCWKRSGSMTSSCYYYSAS 17752

Qy   1088 cggcatacgccaaatacccgtccccttaccattcccgaaatatccgttcaaacttggagc 1147
Db  17753 SCMWMMSSKSCMCCCMMMKRRCACCYSSMSCTSMYRCCWGGMKSYSTCGTTCKKCTGYKK 17812

Qy   1148 agcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaacggcaaaaatg 1207
Db  17813 SRTWYWWMYSWTSKMYMWRAACMMYCYMSTRRMMCMMSMCARKGWCMCTMKGAGATCACA 17872

Qy   1208 tcaaactggcagaccaac 1225
Db  17873 GTAGAGTGAGATAGAAAC 17890
```

```
RESULT  41
AF429315/c
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin;G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN
```

```
  Query Match             2.6%;  Score 48.4;  DB 9;  Length 125020;
  Best Local Similarity   12.6%;  Pred. No. 1.8;
  Matches  124;  Conservative  393;  Mismatches  457;  Indels    7;  Gaps    3;

Qy    837 agatatggcgcaactcaaagactatgccgcagcagccatccgcgattgggcagtccaaaa 896
Db  52150 RGRGMKTGYRGWMWRSMCWSCCYCWKCMCACMSMMAKSCMCKSCYMYMSMMYRGSWSTGC 52091

Qy    897 ccccaatgccgcacaaggcatagaagccgtcagcaatatctttatggcagccatccccat 956
Db  52090 CMCYKYSRRSKSCYKGSWGSYWASMRSYKSCASCWGWSTSYRSKRGGCWSATSYKCMYGS 52031

Qy    957 caaagggattggagctgtccggggaaaatacggcttgggcggcatcacggcacatcctgt 1016
Db  52030 KTRSSTGGKRSACMCWKWCWSSNSR---SMCMSMYAWKCWGKMYWYRYRSRKRYSCYRGS 51974

Qy   1017 caagcggtcgcagatgggcgcgatcgcattgccgaaagggaaatccgccgtcagcgacaa 1076
Db  51973 KCYKSRMAKSMMMRGSMRSSYYMWRMCMWKGRGCRRGAGSAARSYKKGAMKSKGMRGCWG 51914

Qy   1077 ttttgccgatgcggcatacgccaaatacccgtcccccttaccattcccgaaatatccgttc 1136
Db  51913 STYYGKMCGRGSKSMYTGSMKCAATGMKSWSMMCMGRCAMWSKGGSYYRGWGCWCMSYWW 51854

Qy   1137 aaacttggagcagcgttacggcaaagaaaacatcacctcctcaaccgtgccgccgtcaaa 1196
Db  51853 CASYYSRWRGSAWSYWAGGYTSYRKGSW---GWGKMYKGSRARGSMSRWSKRSWWKKSWR 51797

Qy   1197 cggcaaaaatgtcaaactggcagaccaacgccacccgaagacaggcgtaccgtttgacgg 1256
Db  51796 KGRSRASARWKSYSCASRRKSCWRMMRGCSYMTGRSSMGTMSKRKGSRTSMGSKAGTYRS 51737

Qy   1257 taaagggtttccgaattttgagaagcacgtgaaatatgatacgaagctcgatattcaaga 1316
Db  51736 ARRAAMYRSCSKGRRRGKKKSSKGGGRGGKKWMASMTYSKRCCAKSTGRRRGGTKGSSCW 51677

Qy   1317 attatcgggggggcggtatacctaaggctaagcctgtgtttgatgcgaaaccgagatggga 1376
Db  51676 TKRRGWGRTRKAAAWCAKCCWWMYRCTGMCWAKTYYATWYKSWKSRWTTGWRRMGTRRAW 51617

Qy   1377 ggttgataggaagcttaataaattgacaactcgtgagcaggtggagaaaaatgttcagga 1436
Db  51616 MWWWYWWKKYYRARMKCAKTRSMWMASKYSKSYARKCCSMWSAYKKYKSRASRCKGWGKM 51557

Qy   1437 aacgagaagaaggagtcagagtagtcagtttaaagcccatgcgcaacgagaat-gggaaa 1495
Db  51556 AKSCMRRYMMYKWGGGAGKCCSAGGMTGRTGSRWSACCWKMSKKSASKARYAYASTGAAA 51497

Qy   1496 ataaaacagggttagattttaatcattttataggtggtgatatcaataagaaaggcacag 1555
Db  51496 MSMCGTCYMKMCTRRMMATACAWRRWRAWWMSCCMRKCWTSKWSKMGRRMRYCWSWARWM 51437

Qy   1556 taacaggagggcatagtctaacccgtggtgatgtacgggtgatacaacaaacctcggcac 1615
Db  51436 YYAGSYMSKCRSRWGGYKGRSRSRSGMSWRTRGYSYSARCYMSKSAGGAGKCKGAGSYAK 51377

Qy   1616 ctgataaacatggggtttatcaagcgacagtggaaattaaaaagcctgatggaagttggg 1675
```

```
Db  51376 GSAGWRWGSYKWGAWYSCRMSASKKSRMTCYWSCCCTSCKGCYWMYWSWKYGAGMCWYMR 51317

Qy   1676 aggtgaaaacgaaaaaaggtgggaaagtgatgaccaagcacaccatgttcccaaaagatt 1735
Db  51316 WCCTCMARAMWRARMAAMRRARMAGAARAWMSMWMWSMMMMRBHAAAMMARMWAAAKCWW 51257

Qy   1736 gggatgaggctagaattagggctgaagttacttcggcttgggaaagtagaataatgctta 1795
Db  51256 WYSYRKYCWMSMMMASWAAMAMWMMCTAAWYYKKWTRMRRWSKYSSSYAYAWYATMCWYC 51197

Qy   1796 aggataataaatggcagggta 1816
Db  51196 YSSMYAARRMGTYSGGKKRKW 51176


RESULT  42
AC006179/c
LOCUS       AC006179               41755 bp    DNA     linear   HTG 30-AUG-2001
DEFINITION  Homo sapiens chromosome 10 clone LA10NC01_173_E_2 map 10q24.2-24.3,
            *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.
ACCESSION   AC006179
VERSION     AC006179.1  GI:3983522
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 41755)
  AUTHORS   Smith,D.R.
  TITLE     Sequencing of Human Chromosome 10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41755)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1998) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02154, USA
COMMENT     Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing
            Technology. Data may contain low quality seq uence and BAC/Cosmid
            vector sequences.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6416: contig of 6416 bp in length
            *                   gap of unknown length
            *     6417    17234: contig of 10818 bp in length
            *                   gap of unknown length
            *    17235    20214: contig of 2980 bp in length
            *                   gap of unknown length
            *    20215    41755: contig of 21541 bp in length.
FEATURES            Location/Qualifiers
     source         1. .41755
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="10"
                    /map="10q24.2-24.3"
                    /clone="LA10NC01_173_E_2"
BASE COUNT    10256 a  10211 c  10361 g  10914 t     13 others
ORIGIN


  Query Match             2.4%;  Score 45.8;  DB 2;  Length 41755;
  Best Local Similarity   57.2%;  Pred. No. 6.4;
  Matches   83;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy   1414 caggtggagaaaaatgttcaggaaacgagaagaaggagtcagagtagtcagtttaaagcc 1473
Db  36046 CTGGGGCAGAAAAATGTTTGGGCTCCAACCAGAAACAGCTAGAGTCACAGATTTAAAACC 35987

Qy   1474 catgcgcaacgagaatgggaaaataaaacagggttagattttaatcattttataggtggt 1533
Db  35986 CGGGAGTAAAAACATCAGGAAAATCAAGATTGTTCAGAATCATCTACTGGCAGCAGTGAA 35927

Qy   1534 gatatcaataagaaaggcacagtaa 1558
Db  35926 GATAACAGTCTGAGGGGCACACCAA 35902


RESULT  43
AR135147
LOCUS       AR135147               390 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 7 from patent US 6194550.
ACCESSION   AR135147
VERSION     AR135147.1  GI:14124052
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Gold,L., Tuerk,C., Pribnow,D. and Smith,J.Drew.
  TITLE     Systematic polypeptide evolution by reverse translation
  JOURNAL   Patent: US 6194550-A 7 27-FEB-2001;
FEATURES            Location/Qualifiers
     source         1. .390
                    /organism="unknown"
BASE COUNT      125 a    126 c    133 g      6 t
ORIGIN


  Query Match             2.4%;  Score 45.2;  DB 6;  Length 390;
  Best Local Similarity   44.8%;  Pred. No. 4.6;
  Matches  173;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy    309 gcccttcgacaaccatgcctcacattccgattctgacgaagccggtagtcccgttgacgg 368
Db      2 GGCCATGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 61

Qy    369 attcagcctttaccgcatccattgggacggatacgaacaccatcccgccgacggctatga 428
Db     62 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 121

Qy    429 cgggccacagggcggcggctatcccgctcccaaaggcgcgagggatatatacagctacga 488
Db    122 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 181

Qy    489 cataaaaggcgttgcccaaaatatccgcctcaacctgaccgacaaccgcagcaccggaca 548
Db    182 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 241

Qy    549 acggcttgccgaccgtttccacaatgccggcgctatgctgacgcaaggagtaggcgacgg 608
Db    242 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 301

Qy    609 attcaaacgcgccacccgatacagccccgagctggacagatcgggcaatgccgccgaagc 668
Db    302 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 361

Qy    669 cttcaacggcactgcagatatcgtca 694
Db    362 CGACGACGCCATGGTTGCGATGGTCA 387


RESULT  44
AP002824
LOCUS       AP002824               115890 bp    DNA     linear   HTG 28-JUL-2000
DEFINITION  Homo sapiens chromosome 4 clone 56A17 map 4q22-q24, *** SEQUENCING
            IN PROGRESS ***, 2 unordered pieces.
ACCESSION   AP002824
VERSION     AP002824.1  GI:9581848
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens DNA, clone:56A17.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 115890)
  AUTHORS   Tsai,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2000) Shih-Feng Tsai, National Yang-Ming
            University, Institute of Genetics; 155 Li-Rong St. Section 2,
            Peitou, Taipei, Taiwan 11221, Republic of China
            (E-mail:ympetsai@ym.edu.tw, URL:http://genome.ym.edu.tw/,
            Tel:886-2-28267043, Fax:886-2-28264930)
COMMENT     gaps between the contigs are represented as 100 N.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    41185: contig of 41185 bp in length
            *    41186 41285: gap of      100 bp
            *    41286   115890: contig of 74605 bp in length.
FEATURES             Location/Qualifiers
     source          1. .115890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q22-q24"
                     /clone="56A17"
BASE COUNT    35514 a  21174 c  21793 g  37309 t    100 others
ORIGIN
```

```
  Query Match             2.4%;  Score 45.2;  DB 2;  Length 115890;
  Best Local Similarity   51.0%;  Pred. No. 10;
  Matches  107;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy   1658 agcctgatggaagttgggaggtgaaaacgaaaaaaggtgggaaagtgatgaccaagcaca 1717
          ||| |  ||  |   |  |||   | | ||| | ||     | | |||  |  ||  |
Db 112818 AGCTTTCTGAGATAAGACAGGCAGACAAGAATAGAGAAAAAAGAATGAAAAGAAACAAAC 112877

Qy   1718 ccatgttcccaaaagattgggatgaggctagaattagggctgaagttacttcggcttggg 1777
            |   | | ||||    ||||| | |  | |   ||| |||||  |||  |   |||||
Db 112878 AAAAACTGCAAAAAATAAGGGATTATGTAAAACAAAGGACTGAACCTACAACTAATTGGG 112937

Qy   1778 aaagtagaataatgcttaaggataataaatggcagggtacaagtaaatcgggtattaaaa 1837
            |   |||  | |     ||  | ||      ||| | |  ||  || ||   | |
Db 112938 GTACCTGAAAGAGACAGGGAGAACAGAACCAAGTGGGCAAACATACTTCAGGATATCATC 112997

Qy   1838 tagaaggatttaccgaacctaatagaacag 1867
          ||| || | || || ||||||  || ||||
Db 112998 TAGGAGAACTTCCCCAACCTAGAAGGACAG 113027
```

```
RESULT  45
AC096727/c
LOCUS       AC096727              157078 bp    DNA     linear   PRI 03-JAN-2002
DEFINITION  Homo sapiens chromosome 4 clone RP11-35D5, complete sequence.
ACCESSION   AC096727 AC012571
VERSION     AC096727.3  GI:16647281
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157078)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157078)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 157078)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 157078)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 4, 2001 this sequence version replaced gi:16303528.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0035D05
            Drafting center: WIBR
            ------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .157078
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-35D5"
BASE COUNT    49522 a  29310 c  28978 g  49268 t
ORIGIN
```

```
  Query Match             2.4%;  Score 45.2;  DB 9;  Length 157078;
  Best Local Similarity   51.0%;  Pred. No. 11;
  Matches  107;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy   1658 agcctgatggaagttgggaggtgaaaacgaaaaaaggtgggaaagtgatgaccaagcaca 1717
          ||| |  ||  |   |  |||   | | ||| | ||     | | |||  |  ||  |
Db  39157 AGCTTTCTGAGATAAGACAGGCAGACAAGAATAGAGAAAAAAGAATGAAAAGAAATAAAC 39098

Qy   1718 ccatgttcccaaaagattgggatgaggctagaattagggctgaagttacttcggcttggg 1777
            |   | | ||||    ||||| | |  | |   ||| |||||  |||  |   |||||
Db  39097 AAAAACTGCAAAAAATAAGGGATTATGTAAAACAAAGGACTGAACCTACAACTAATTGGG 39038

Qy   1778 aaagtagaataatgcttaaggataataaatggcagggtacaagtaaatcgggtattaaaa 1837
            |   |||  | |     ||  | ||      ||| | |  ||  || ||   | |
Db  39037 GTACCTGAAAGAGACAGGGAGAACAGAACCAAGTGGGCAAACATACTTCAGGATATCATC 38978

Qy   1838 tagaaggatttaccgaacctaatagaacag 1867
          ||| || | || || ||||||  || ||||
Db  38977 TAGGAGAACTTCCCCAACCTAGAAGGACAG 38948
```

```
Search completed: June 30, 2002, 08:42:47
Job time: 55585 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 17:16:22 ; Search time 27321.5 Seconds
(without alignments)
1029.420 Million cell updates/sec

Title: US-09-303-518D-125
Perfect score: 1344
Sequence: 1 atgattaaaatcaaaaaagg.....ccattgagaaggaaggctga 1344

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1344	100.0	1344	6	A96092	A96092 Sequence 12
2	c	1344	100.0	12273	1	AE002412	AE002412 Neisseria
3	c	1344	100.0	349980	6	AX044029	AX044029 Sequence
4	c	1244.8	92.6	311321	1	NMA3Z2491	AL162754 Neisseria
5		1229.4	91.5	1344	6	A96094	A96094 Sequence 12
6		1212.8	90.2	1344	6	A96098	A96098 Sequence 13
7		560	41.7	10563	1	AE006171	AE006171 Pasteurel
8		556.6	41.4	1558	1	APU24492	U24492 Actinobacil
9		524.6	39.0	13299	1	U32702	U32702 Haemophilus
10		505.6	37.6	6048	1	AF165980	AF165980 Vibrio ha
11		489.6	36.4	5916	1	AF117331	AF117331 Vibrio ch
12	c	489.6	36.4	13382	1	AE004300	AE004300 Vibrio ch
13	c	477.2	35.5	16903	1	AE004724	AE004724 Pseudomon
14		475.2	35.4	5078	1	VANQRBOL	Z37111 V.alginolyt
15		473.6	35.2	474	6	A96090	A96090 Sequence 12
16		473.6	35.2	6908	1	AB008030	AB008030 Vibrio al
17	c	326.4	24.3	96109	6	AX067460	AX067460 Sequence
18		113.6	8.5	1356	1	AF144076	AF144076 Porphyrom
19	c	100.6	7.5	4926	1	HIU20229	U20229 Haemophilus
20		47.4	3.5	125020	9	AF429315	AF429315 Homo sapi
21	c	46	3.4	36200	1	BACGNTZA	D78193 Bacillus su
22	c	46	3.4	215534	1	BSUB0021	Z99124 Bacillus su
23	c	41.4	3.1	109730	2	OSJN00193	AL662992 Oryza sat
24		41.4	3.1	142805	2	OSJN00146	AL662935 Oryza sat
25		40.2	3.0	151329	2	AP003456	AP003456 Oryza sat
26		39.8	3.0	216050	1	AL646076	AL646076 Ralstonia
27	c	39.2	2.9	24268	1	AB070947	AB070947 Streptomy
28	c	39.2	2.9	187948	9	AP003402	AP003402 Homo sapi
29		38.2	2.8	37245	1	SC5F2A	AL049587 Streptomy
30		38.2	2.8	38734	1	SC10A5	AL021529 Streptomy
31		38	2.8	975	1	AF103943	AF103943 Streptomy
32		37.8	2.8	1694	8	ATY14325	Y14325 Arabidopsis
33		37.8	2.8	11351	1	AE004619	AE004619 Pseudomon
34	c	37.8	2.8	11660	1	AE004956	AE004956 Pseudomon
35	c	37.8	2.8	103573	8	AC005499	AC005499 Arabidops
36	c	37.6	2.8	10944	1	AE002563	AE002563 Neisseria
37	c	37.6	2.8	172325	6	AX044035	AX044035 Sequence
38		37.6	2.8	198050	1	AL646061	AL646061 Ralstonia
39	c	37.6	2.8	340806	1	NMA1Z2491	AL162752 Neisseria
40		37.4	2.8	4499	8	AF324244	AF324244 Phaseolus
41	c	37.4	2.8	7365	1	STMSVIDEOX	L37334 Streptomyce
42		37.4	2.8	39250	1	SVI011500	AJ011500 Streptomy
43	c	37.4	2.8	125020	9	AF429315	AF429315 Homo sapi
44		37.4	2.8	154330	2	AC104429	AC104429 Oryza sat
45		37.2	2.8	3867	1	AY033407	AY033407 Myxococcu
46	c	37.2	2.8	4130	8	AF190427	AF190427 Colletotr
47		37	2.8	2557	3	AY071018	AY071018 Drosophil
48	c	37	2.8	5021	1	AX024382	AX024382 Sequence
49	c	37	2.8	5021	6	AX024275	AX024275 Sequence
50		37	2.8	37750	1	MLCB1913	AL022118 Mycobacte
51		37	2.8	89791	3	AC005268	AC005268 Drosophil
52	c	37	2.8	101085	2	AC018341	AC018341 Drosophil
53		37	2.8	164035	3	AC007888	AC007888 Drosophil
54		37	2.8	213050	1	AL646079	AL646079 Ralstonia
55		37	2.8	269203	1	MLEPRTN10	AL583926 Mycobacte
56	c	37	2.8	305020	3	AE003452	AE003452 Drosophil
57		36.8	2.7	10500	1	AE005062	AE005062 Halobacte
58		36.8	2.7	17589	1	AE004862	AE004862 Pseudomon
59	c	36.4	2.7	1725	6	AX073998	AX073998 Sequence
60	c	36.4	2.7	2076	6	AX073995	AX073995 Sequence
61	c	36.4	2.7	2109	6	AX350975	AX350975 Sequence
62		36.4	2.7	10486	1	AE001656	AE001656 Chlamydia
63	c	36.4	2.7	15680	1	AE002164	AE002164 Chlamydop
64	c	36.4	2.7	33801	1	SCC77	AL136503 Streptomy
65		36.4	2.7	38634	1	SC26G5	AL359152 Streptomy
66		36.4	2.7	300550	1	AP002547	AP002547 Chlamydop
67		36.2	2.7	4623	1	AF419157	AF419157 Aeromonas
68		36	2.7	760	4	AY011741	AY011741 Canis fam
69		36	2.7	975	4	AY011738	AY011738 Felis cat
70		36	2.7	975	4	AY011739	AY011739 Leopardus
71		36	2.7	975	4	AY011740	AY011740 Panthera
72		36	2.7	11526	1	AF228498	AF228498 Escherich
73		36	2.7	349116	1	AP003003	AP003003 Mesorhizo

	74	35.8	2.7	1421	1	STMISAA M31716 S.clavulige
	75	35.8	2.7	3855	13	M23555 M23555 Figure 2. N
c	76	35.8	2.7	10029	1	AE009661 AE009661 Brucella
	77	35.8	2.7	13579	1	AE005061 AE005061 Halobacte
c	78	35.8	2.7	50361	1	AF454951 AF454951 Brucella
	79	35.8	2.7	154918	9	HSDJ67A8 AL121957 Human DNA
c	80	35.8	2.7	174301	2	AP003523 AP003523 Oryza sat
	81	35.8	2.7	181680	2	AL671934 AL671934 Homo sapi
	82	35.6	2.6	3630	1	MTBMTK L33465 Methylobact
	83	35.6	2.6	4081	14	AF274291 AF274291 Culex nig
	84	35.6	2.6	9337	1	AE005717 AE005717 Caulobact
	85	35.6	2.6	15504	1	AE006995 AE006995 Mycobacte
	86	35.6	2.6	22550	1	MTCY22G8 Z95585 Mycobacteri
	87	35.6	2.6	108252	14	AF403738 AF403738 Culex nig
	88	35.6	2.6	123580	1	AF263912 AF263912 Streptomy
	89	35.6	2.6	125401	6	AX211739 AX211739 Sequence
	90	35.4	2.6	296	6	AX260544 AX260544 Sequence
c	91	35.4	2.6	763	6	AX019436 AX019436 Sequence
c	92	35.4	2.6	1598	6	AX019435 AX019435 Sequence
	93	35.4	2.6	9810	1	AF018073 AF018073 Rhodobact
	94	35.4	2.6	19981	1	AE008823 AE008823 Salmonell
c	95	35.4	2.6	30132	1	SC10G8 AL158057 Streptomy
	96	35.4	2.6	40571	1	MSGDNAB L39923 Mycobacteri
c	97	35.4	2.6	58918	8	NCB13I18 AL390189 Neurospor
	98	35.4	2.6	149699	8	AP002486 AP002486 Oryza sat
	99	35.4	2.6	158133	8	AP002816 AP002816 Oryza sat
	100	35.2	2.6	6149	1	AF337647 AF337647 Xanthomon

ALIGNMENTS

```
RESULT   1
A96092
LOCUS       A96092                  1344 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 125 from Patent WO9924578.
ACCESSION   A96092
VERSION     A96092.1  GI:6779938
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1344)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 125 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1344
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      329 a    366 c    365 g    284 t
ORIGIN
```

Query Match 100.0%; Score 1344; DB 6; Length 1344;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 60

Qy      61 tacgacggcccggccattaccgaagtcgcgttgcttgcttggcgaagaatatgccggtatgcgc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy     121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy     181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 240

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy     301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 360

Qy     361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 420

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 480

Qy     481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 540

Qy     541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 600

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 660

Qy     661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 720

Qy     721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 780

Qy     781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 900

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 960

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy    1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 1080

Qy    1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 1140

Qy    1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 1200

Qy    1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 1320

Qy    1321 gaaaccattgagaaggaaggctga 1344
           |||||||||||||||||||||||
Db    1321 GAAACCATTGAGAAGGAAGGCTGA 1344


RESULT   2
AE002412/c
LOCUS       AE002412              12273 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 54 of 206 of
            the complete genome.
ACCESSION   AE002412 AE002098
VERSION     AE002412.1  GI:7225783
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 12273)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 12273)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12273
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            97. .435
                     /gene="NMB0557"
     CDS             97. .435
                     /gene="NMB0557"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40985.1"
                     /db_xref="GI:7225784"
                     /translation="MSDESPIIFTDSCCAKVADLIAEENNPDLKLRVFVNGGGCSGFQ
                     YGFTFDEIKNDDDFEIEKNGLVFLVDPMSYQYLVGAEIDYTESLQGSQFVIRNPNAET
                     TCGCGSSFSV"
     gene            595. .1059
                     /gene="NMB0558"
     CDS             595. .1059
                     /gene="NMB0558"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40986.1"
                     /db_xref="GI:7225785"
                     /translation="MIEMSKDYRNDLYDVYVSYPPQVDRGLIRECLKENLGEEKAEGL
                     IESLDSKPQVLVEEKCTWAKREELHDYFSYLGLDIITRRYMELETVVPPEEGEGEGEG
                     ADGEMPEYLELHGGREDDISAPSQPEPPSRNIKLLVFGLLIAFLGYLLGKIF"
     gene            1094. .2605
                     /gene="NMB0559"
     CDS             1094. .2605
                     /gene="NMB0559"
                     /note="similar to GB:M87049 SP:P27853 PID:148233 GB:U00096
                     PID:2367309 percent identity: 74.73; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ubiquinone biosynthesis protein AarF"
                     /protein_id="AAF40987.1"
                     /db_xref="GI:7225786"
                     /translation="MKWLKRLTVIVGTFYRYRLAGLCASLMGSGWICALLKMMPQSSK
                     LKNEPPAVRLRLALESLGPIFIKFGQVLSTRPDLIPHDYAVELAKLQDKVPPFDARLS
                     REQIEKSLGQSIEKLYAEFETEPIASASIAQVHKARLHSGEQVAVKVLRPNLLPVIEQ
                     DLSLMRFGAGWVERLFADGKRLKPREVVAEFDKYLHDELDLMREAANASQLGRNFQNS
                     DMLIVPKVFYDYCTSDVLTIEWMDGTPVSDIAKLKADGIDLHKLADYGVEIFFTQVFR
                     DGFFHADMHPGNILVAADNRYIALDFGIVGTLTDYDKRYLAINFLAFFNRDYRRVATA
                     HIESGWVPADTRAEELEAAVRAVCEPVFNKPISQISFGLVLMRLFEVSRRFNVEIQPQ
                     LVLLQKTLLNIEGLGRQLDPDLDLWKTAKPFLVKWMNGQVGPKALWRNLKNEAPDWAQ
                     IIPSLPRKISALIDENRQQEMRDAYIHLVKVQQRQSLWLAVIAVVLLLILLLK"
     gene            complement(2962. .3780)
                     /gene="NMB0560"
     CDS             complement(2962. .3780)
                     /gene="NMB0560"
                     /note="similar to GB:L42023 SP:P43886 PID:1005429
                     PID:1220683 PID:1204857 percent identity: 72.97;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="serine acetyltransferase"
                     /protein_id="AAF40988.1"
                     /db_xref="GI:7225787"
                     /translation="MKKDHLNTTGFDLWHTIREETAAAVSAEPMLASFLHQTVLRHES
                     LGSVLAYHLSSKLGSPIMDVRALFEIYQQALGSDTQIGKCVEADLKAIYERDPACDEY
                     SLPLLYFKGFHAIQAHRINHRLYLDGRKTLAYFLQNRMSEVFGVDIHPAARFGYGLML
                     DHATGFVAGETAVLGNNISILHGVTLGGSGKEGGDRHPKIGDGVMIGANASILGNIRI
                     GSNAKIGAGSVVVSDVPPSITVVGVPAKPVARSLKTPSADMDQNIQFTEIDFMI"
     gene            3964. .4542
                     /gene="NMB0561"
     CDS             3964. .4542
                     /gene="NMB0561"
                     /note="similar to GP:3777486 percent identity: 61.03;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="grpE protein"
                     /protein_id="AAF40989.1"
                     /db_xref="GI:7225788"
                     /translation="MSEQTQQQNSEEAVENVEAVETVETVGNADGVQEQAAAEPAYED
                     LQARIAELEAQLKDEQLRALANEQNLRRRHQQEIADTHKFAGQKFAVEMLPVKDYLEM
                     ALLDQSGNFDALKMGVQMTLNELQKAFDATQIKEINPKAGDKLDPNIHQAMQAVASEQ
                     EPNTVVGVMKKGYTLSDRVLRPAMVTVAQKEA"
     gene            complement(4819. .5034)
                     /gene="NMB0562"
     CDS             complement(4819. .5034)
                     /gene="NMB0562"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40990.1"
                     /db_xref="GI:7225789"
                     /translation="MKTLLLTFGIFLTVIIGMAVGYIFSKRSIKGSCGGITALGMKKM
```

```
                CDCDTPCDTLQKKLDEEKQAGGIRVDR"
     gene       complement(5060. .6115)
                /gene="NMB0563"
     CDS        complement(5060. .6115)
                /gene="NMB0563"
                /note="similar to GB:L42023 SP:P44550 PID:1003244
                PID:1222087 PID:1204428 percent identity: 74.70;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="thiamine biosynthesis lipoprotein ApbE"
                /protein_id="AAF40991.1"
                /db_xref="GI:7225790"
                /translation="MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTT
                YTVKYLSNNRDKLPSPAEIQKRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRIS
                SDFAHVTAEAVRLNRLTHGALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGI
                DKIILKQGKDYASLSKTHPKAYLDLSSIAKGFGVDKVAGELEKYGIQNYLVEIGGELH
                GKGKNARGEPWRIGIEQPNIVQGGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHI
                INPNNKRPISHNLASISVVADSAMTADGLSTGLFVLGETEALKLAEREKLAVFLIVRD
                KGGYRTAMSSEFEKLLR"
     gene       complement(6268. .7485)
                /gene="NMB0564"
     CDS        complement(6268. .7485)
                /gene="NMB0564"
                /note="similar to GB:L42023 PID:1003242 PID:1222086
                PID:1204427 PID:1573127 percent identity: 92.82;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="Na(+)-translocating NADH-quinone reductase,
                subunit F"
                /protein_id="AAF40992.1"
                /db_xref="GI:7225791"
                /translation="MEILGIVMFTVIVLVLALMILFAKSKLVSEGDITIKVNGEKEL
                TMPAGGKLLGALANEGIFIPSACGGGGSCGQCRVVVKSGGGDILPTELSHISKREARE
                GCRLSCQVNVKTDMDIEVPEEVFGVKKWECTVISNDNKATFIKELKLAIPEGEEVPFR
                AGGYIQIEAPPHTVAYKDFDIPKEYHEDWDKYNLWQYVSKVDEPILRAYSMASYPEEK
                GIIMLNVRIATPPPRVPDAPPGQMSSYIWSLKPGDKVTISGPFGEFFAKDTDAEMVFI
                GGGAGMAPMRSHIFDQLKRLNSKRKITFWYGARSKREMFYVEDFDQLAAEFPNFTWHV
                ALSDPLPEDNWDGYTGFIHNVVYENHLKNHEAPEDCEFYMCGPPIMNQSVIKMLKDLG
                VEDENILLDDFGG"
     gene       complement(7499. .8092)
                /gene="NMB0565"
     CDS        complement(7499. .8092)
                /gene="NMB0565"
                /note="similar to PID:1573126 percent identity: 91.92;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="Na(+)-translocating NADH-quinone reductase,
                subunit E"
                /protein_id="AAF40993.1"
                /db_xref="GI:7225792"
                /translation="MEHYLSLFIKSVFIENMALSFFLGMCTFLAVSKKVSTAFGLGVA
                VIFVLGLSVPVNQLVYSLLKDGAIAEGVDLTFLKFITFIGVIAALVQILEMFLDKFVP
                ALYNALGIYLPLITVNCAIFGAVSFMAQREYNFGESVVYGFGAGLGWMLAIVALAGIT
                EKMKYSDAPKGLKGLGITFIAAGLMAMAFMSFSGIQL"
     gene       complement(8096. .8722)
                /gene="NMB0566"
     CDS        complement(8096. .8722)
                /gene="NMB0566"
                /note="similar to GB:L42023 SP:P43958 PID:1003237
                PID:1222083 PID:1204425 percent identity: 89.42;

  Query Match             100.0%;  Score 1344;  DB 1;  Length 12273;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1344;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12069 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 12010

Qy      61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12009 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 11950

Qy     121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11949 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 11890

Qy     181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11889 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 11830

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11829 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 11770

Qy     301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11769 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 11710

Qy     361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11709 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 11650

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11649 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 11590

Qy     481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11589 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 11530

Qy     541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11529 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 11470

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11469 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 11410

Qy     661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11409 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 11350

Qy     721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11349 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 11290

Qy     781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11289 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 11230

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11229 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 11170

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11169 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 11110

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11109 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 11050

Qy    1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11049 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 10990

Qy    1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10989 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 10930

Qy    1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  10929 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 10870

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10869 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 10810

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10809 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 10750

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||||||||||||||||||||
Db  10749 GAAACCATTGAGAAGGAAGGCTGA 10726


RESULT   3
AX044029/c
LOCUS       AX044029               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 108 from Patent WO0066791.
ACCESSION   AX044029
VERSION     AX044029.1  GI:11342913
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 108 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    84410 a  84863 c  94187 g  86520 t
ORIGIN


  Query Match           100.0%;  Score 1344;  DB 6;  Length 349980;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1344;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297909 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 297850

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297849 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 297790

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297789 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 297730

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297729 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 297670

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297669 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 297610

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297609 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 297550

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297549 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 297490

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297489 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 297430

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297429 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 297370

Qy    541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297369 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 297310

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297309 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 297250

Qy    661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297249 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 297190

Qy    721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297189 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 297130

Qy    781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297129 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 297070

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297069 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 297010

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297009 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 296950

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296949 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 296890

Qy   1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296889 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 296830

Qy   1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296829 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 296770

Qy   1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296769 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 296710

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296709 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 296650

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296649 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 296590

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||||||||||||||||||||
Db 296589 GAAACCATTGAGAAGGAAGGCTGA 296566


RESULT   4
```

```
NMA3Z2491/c
LOCUS       NMA3Z2491             311321 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 3/7.
ACCESSION   AL162754 AL157959
VERSION     AL162754.2  GI:7379424
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 311321)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 311321)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .311321
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     RBS             115. .120
     gene            129. .587
                     /gene="NMA0698"
     CDS             129. .587
                     /gene="NMA0698"
                     /note="NMA0698, possible periplasmic protein, len: 152 aa;
                     unknown, contains a probable N-terminal signal sequence
                     and lies within a region of unusually low GC content"
                     /codon_start=1
                     /transl_table=11
                     /product="putative periplasmic protein"
                     /protein_id="CAB83985.1"
                     /db_xref="GI:7379425"
                     /db_xref="SPTREMBL:Q9JQQ2"
                     /translation="MLTKLKILLFLFLFVFVLAINLLFFFFSSDIESFGNYQFEYVYD
                     KGWPANYILVMKDGNEGNFDKIISGLVLEYYKEDDNIYFSYIDGQGFASDSCYYKPEI
                     LYGKIILNKNHIININSMEKNNFLSEDKIMKGTRNWLADPKNKCNIQTLD"
     gene            742. .1077
                     /gene="NMA0699"
     CDS             742. .1077
                     /gene="NMA0699"
                     /note="NMA0699, pseudogene, probable ABC transporter
                     protein, len: 336 bp; similar to C-termini of many ABC
                     transporters e.g. TR:Q46973 (EMBL:U47048), mtfB,
                     Escherichia coli microcin transport protein (707 aa),
                     fasta scores; E(): 2.6e-19, 53.2% identity in 111 aa
                     overlap. Contains PS00211 ABC transporters family
                     signature. NMA0686 may be the remainder of this
                     pseudogene"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="putative ABC transporter protein (pseudogene)"
     misc_feature    868. .912
                     /gene="NMA0699"
                     /note="PS00211 ABC transporters family signature"
     gene            complement(1101. .2327)
                     /gene="NMA0700"
     CDS             complement(1101. .2327)
                     /gene="NMA0700"
                     /note="NMA0700, possible ribonuclease BN, len: 408 aa;
                     similar to SW:RBN_ECOLI (EMBL:L19201), rbn, Escherichia
                     coli ribonuclease BN (EC 3.1.-.-), fasta scores; E():
                     2.3e-28, 34.3% identity in 268 aa overlap. Longer than rbn
                     at the C-terminus. Also similar to TR:O85449
                     (EMBL:AF067083) Vitreoscilla sp. hypothetical protein (376
                     aa), fasta scores; E(): 0, 45.0% identity in 333 aa
                     overlap. Contains a region similar to NMA0677, fasta
                     scores; E(): 4.3e-08, 64.4% identity in 45 aa overlap.
                     Contains hydrophobic, possible membrane-spanning regions,
                     as does rbn"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribonuclease BN"
                     /protein_id="CAB83987.1"
                     /db_xref="GI:7379426"
                     /db_xref="SPTREMBL:Q9JVT9"
                     /translation="MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLL
                     ALVPVLTVMVAVASIFPVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTA
                     IGSVMLVVTSLMLIRTIDNTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFM
                     VGSVQDAALASGAPQWSGALRTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATA
                     FCLETARSLFTWYMGNFDGYRSIYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQ
                     GEAFRRGFDSRGRFDDVLKILLLLDAAQKEGKALPVQEFRRHINMGYDELGELLEKLA
                     RHGYIYSGRQGWVLKTGADSIELNELFKLFVYRPLPVERDHVNQAVDAVMMPCLQTLN
                     MTLAEFDAQAKKQQQS"
     misc_feature    1376. .1385
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(2417. .2426)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2444. .2453
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            2486. .2743
                     /gene="NMA0701"
     CDS             2486. .2743
                     /gene="NMA0701"
                     /note="NMA0701, possible pseudogene, len: 258 bp; shows
                     weak similarity to part of SW:WRBA_ECOLI (EMBL:M99166),
                     wrbA, Escherichia coli Trp repressor binding protein (197
                     aa), fasta scores; E(): 0.064, 29.3% identity in 82 aa
                     overlap. Also similar to part of TR:O85450 (EMBL:AF067083)
                     Vitreoscilla sp. Trp repressor binding protein (fragment)
                     (124 aa), fasta scores; E(): 6.6e-14, 56.6% identity in 83
                     aa overlap"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="hypothetical protein NMA0701 (pseudogene)"
     misc_feature    complement(2755. .2764)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2778. .2787
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             2861. .2864
     gene            2871. .3530
                     /gene="NMA0702"
     CDS             2871. .3530
                     /gene="NMA0702"
                     /note="NMA0702, len: 219 aa; similar to TR:Q52925
                     (EMBL:Z50189), exsB, Rhizobium meliloti putative regulator
                     of succinoglycan biosynthesis (not a transcriptional
                     regulator) (234 aa), fasta scores; E(): 3.1e-15, 32.7%
                     identity in 205 aa overlap. Also similar to many
                     hbacterial ypothetical proteins e.g. SW:YBAX_HAEIN
                     (EMBL:U32798), HI1191, Haemophilus influenzae hypothetical
```

```
                     protein (196 aa), fasta scores; E(): 0, 79.2% identity in
                     173 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0702"
                     /protein_id="CAB83989.1"
                     /db_xref="GI:7379427"
                     /db_xref="SPTREMBL:Q9JVT8"
                     /translation="MSNQKALVIFSGGQDSTTCLIQAIQTYGRENVQAITFQYGQRHA
                     VELERARWIAQDLGVKQTVLDLSLMRQITHNALMDDTAAIETAENGVPNTFVDGRNAL
                     FLLYAAIYAKGQGIRHIIAGVCETDFSGYPDCRDVFVKSMNVTLNLAMDYDFQIHTPL
                     MYLTKAQTWALADEMGALDYIREQTHTCYNGIVGGCRECPSCILRERGLAEYLESKKA
                     V"
misc_feature         3522. .3531
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
RBS                  3547. .3550
gene                 3558. .4076
                     /gene="NMA0703"
CDS                  3558. .4076
                     /gene="NMA0703"
                     /note="NMA0703, len: 172 aa; unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0703"
                     /protein_id="CAB83990.1"
                     /db_xref="GI:7379428"
                     /db_xref="SPTREMBL:Q9JVT7"
                     /translation="MPKLHMFYLGGNAGRSNIEVHDIQFAVCDDYREAVPALKAAWFG
                     DADKIHIDGWQVVEWADGYDIAVSETPKTKMPSENAPRLYFANVGGYRAGQLAEAHAF
                     GLFAAATPAEAKQKALQTLLTDSYVQQHKDNLKDVDNLLALDRIGNFHIRLTPNPHGK
                     PAEIGFQGYLPI"
misc_feature         3782. .3791
                     /gene="NMA0703"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
gene                 4083. .4505
                     /gene="NMA0704"
misc_feature         4083. .4502
                     /gene="NMA0704"
                     /note="Pfam match to entry PF01242 PTPS, 6-pyruvoyl
                     tetrahydropterin synthase, score 10.80, E-value 1.7e-06"
CDS                  4083. .4505
                     /gene="NMA0704"
                     /note="NMA0704, len: 140 aa; similar to many hypothetical
                     proteins e.g. SW:YB90_HAEIN (EMBL:U32798), HI1190,
                     Haemophilus influenzae hypothetical protein (141 aa),
                     fasta scores; E(): 5.1e-32, 58.0% identity in 138 aa
                     overlap. Shows very weak similarity to eukaryotic
                     6-pyruvoyl-tetrahydropterin synthases e.g. SW:PTPS_RAT
                     (EMBL:M77850), pts, Rattus norvegicus 6-pyruvoyl
                     tetrahydrobiopterin synthase precursor (EC 4.6.1.10) (144
                     aa), fasta scores; E(): 0.0062, 27.3% identity in 154 aa
                     overlap. Contains Pfam match to entry PF01242 PTPS,
                     6-pyruvoyl tetrahydropterin synthase"

  Query Match           92.6%;  Score 1244.8;  DB 1;  Length 311321;
  Best Local Similarity  95.4%;  Pred. No. 1.9e-294;
  Matches 1282;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| | ||
Db   55476 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 55417

Qy      61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           || |||||  || | ||||||||||||||||||||||||||||||||||||||||||||
Db   55416 TATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 55357

Qy     121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55356 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 55297

Qy     181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
           |||||||||||||||||||||||||||||||||||||| ||||||| |||||||||||||| || ||
Db   55296 AAAAAGAATCCGGGCGTGGTGTTTACCGCGCCGGTTTCAGGCAAAATCGCCGCCATCCAT 55237

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           || |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   55236 CGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 55177

Qy     301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
           ||||| |||||||||| || |||||| ||||||||||||||||||||||||||||||||
Db   55176 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 55117

Qy     361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
           || |||||||||||||||||||||||||||||||| ||||||||||||||||||||| |||
Db   55116 AATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTACCCGTCCGTTCAGCAAAATCCCT 55057

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55056 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCG 54997

Qy     481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
           || ||||||  |||  | |||||||||||||||||||||||| ||| ||  || |||||
Db   54996 GCAGACCCTGTGGTTGTGATCAAAGAAGCCGCCGAGGATTTCAGACGAGGTTTGCTGGTA 54937

Qy     541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
           |||||||||||||||| || |||||||||| ||||||||||||||||||||||||||||
Db   54936 TTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 54877

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   54876 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTG 54817

Qy     661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
           |||||||||||||||||||| |||||||||| || || |||||||| |||||||||||
Db   54816 AGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATC 54757

Qy     721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
           |||||||||||||||||| |||| || ||||||||||||||||||||||||||||||||
Db   54756 AATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 54697

Qy     781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
           |||||||||||  | |||||||||||||||||||||| ||||||||||||||||||||
Db   54696 CGCGTGATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTG 54637

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   54636 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 54577

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54576 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 54517

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54516 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 54457

Qy    1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   54456 CCGCAGCCGGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAA 54397

Qy    1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
           |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db   54396 CTCTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACT 54337

Qy    1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
           |||||||||||  ||||| | || |||||||| ||||||||||||||||||||||||||||
Db   54336 TACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 54277

Qy    1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   54276 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 54217

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||| || ||||||
```

```
Db  54216 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGTAAGGTGCTG 54157

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||||||||||||||||||||||
Db  54156 GAAACCATTGAGAAGGAAGGCTGA 54133


RESULT   5
A96094
LOCUS       A96094                   1344 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 127 from Patent WO9924578.
ACCESSION   A96094
VERSION     A96094.1  GI:6779939
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1344)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 127 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1344
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      323 a    347 c    368 g    289 t     17 others
ORIGIN


  Query Match             91.5%;  Score 1229.4;  DB 6;  Length 1344;
  Best Local Similarity   94.2%;  Pred. No. 8.6e-291;
  Matches 1266;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| || ||
Db      1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 60

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
          || ||||| || | ||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCTNGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          |||||| |||||||||||||||||| ||||| | |||||||||||||||||| || ||
Db    181 AAAAAGNATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCAT 240

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          || |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    241 CGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
          ||||| |||||||||||| || |||||| |||||||||||||||||| ||| |  |  ||
Db    301 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGAANTNNGNNGC 360

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
          || |||||||||||||||||||||||||||||||| | |||||||||||||||||| |||
Db    361 AATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCT 420

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGCG 480

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          || ||||||  |||  | |||||||||||||| ||| ||||||| |||  |  || |||||
Db    481 GCAGACCCTGTGGTTGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTA 540

Qy    541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
          |||||||||||||||| || |||||||||| |||||||||||||||||||||||||||||
```

```
Db    541 TTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 600

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTG 660

Qy    661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
          |||||||||||||||||||||||| |||||||||| || || |||||||| |||||||||
Db    661 AGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATC 720

Qy    721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
          |||||||||||||||||| |||| || ||||||||||||||||||||||||||||||||||
Db    721 AATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 780

Qy    781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          |||||||||||  | |||||||||||||||||||||| ||||||||||||||||||||||
Db    781 CGCGTGATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTG 840

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    841 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 960

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy   1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   1021 CCGCAGCCGGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAA 1080

Qy   1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
          |||||||||||||  |||||||||||||| ||||||||||||||||||||||||||||||
Db   1081 CTCTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACT 1140

Qy   1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
          |||||||||||  |||||  | || |||||||| ||||||||||||||||||||||||||
Db   1141 TACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 1200

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          |||||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db   1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCCCGCTGTTGCGTAAGGTGCTG 1320

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||| ||||||||||||||||||
Db   1321 GAAACCNTTGAGAAGGAAGGCTGA 1344


RESULT   6
A96098
LOCUS       A96098                   1344 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 131 from Patent WO9924578.
ACCESSION   A96098
VERSION     A96098.1  GI:6779941
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1344)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 131 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1344
```

```
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      334 a    363 c    362 g    285 t
ORIGIN


  Query Match           90.2%;  Score 1212.8;  DB 6;  Length 1344;
  Best Local Similarity  93.9%;  Pred. No. 1e-286;
  Matches 1262;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||| | ||
Db      1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 60

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
          || |||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||
Db     61 TATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGC 120

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          |||||||||||| |||||||||| || ||||||||||||||||||||||||||||||||||
Db    121 CCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          |||||||||||||||| || |||||||||||||||||||||||||||||||| ||||||
Db    181 AAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCAC 240

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
          ||||| |||||||||| |||||||||||||||||| || ||| ||||| ||||||||||
Db    301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGC 360

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
          |||||||| ||||| || || |||||||||| |||||||||||||||||||||||| |||
Db    361 AACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCT 420

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 480

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          |||||||||||||||| |||||||||||||||| || ||||||||||||||||||||||
Db    481 GCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTA 540

Qy    541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
          ||||||||  |||||||||| |||||||||| ||||| ||||| ||||||||||||||||
Db    541 TTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
          ||||||||||||||||| |||||||||||||||| |||||||||||||||||||||| |||
Db    601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTG 660

Qy    661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    661 AGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATC 720

Qy    721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
          |||||||||| || ||| | || || |||||||| | ||||||||||||||| ||||||
Db    721 AATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAG 780

Qy    781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          |||||| |||||| | || ||    ||||||||||||||||||||||||||||||||||||
Db    781 CGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
          |||||||| || || ||| |||| || ||||||||||||||| | ||||||||||||||||
Db    841 GGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          |||||||||||||||||| |||||| |||||||||||| ||||||||||||||||||||||
Db    901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTAC 960

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy   1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
          |||||||||||||||||||||||||||||| || || |||||||||||||| |||||||||
Db   1021 CCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAA 1080

Qy   1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
          |||||||||||||  ||||||||||||||||||||||||||||||| ||||| || |||
Db   1081 CTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACT 1140

Qy   1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
          || |||||||| ||||| ||||| |||||||| ||| ||||||||||||||||||||||||
Db   1141 TATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 1320

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||||||||||||||||||||
Db   1321 GAAACCATTGAGAAGGAAGGCTGA 1344


RESULT   7
AE006171
LOCUS       AE006171               10563 bp    DNA     linear   BCT 08-MAR-2001
DEFINITION  Pasteurella multocida PM70 section 138 of 204 of the complete
            genome.
ACCESSION   AE006171 AE004439
VERSION     AE006171.1  GI:12721691
KEYWORDS    .
SOURCE      Pasteurella multocida.
  ORGANISM  Pasteurella multocida
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Pasteurella.
REFERENCE   1  (bases 1 to 10563)
  AUTHORS   May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
            Kapur,V.
  TITLE     Complete genomic sequence of Pasteurella multocida,Pm70
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
  MEDLINE   21145866
REFERENCE   2  (bases 1 to 10563)
  AUTHORS   Zhang,Q. and Kapur,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
            University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
            55108, USA
FEATURES             Location/Qualifiers
     source          1. .10563
                     /organism="Pasteurella multocida"
                     /strain="PM70"
                     /db_xref="taxon:747"
     gene            94. .1605
                     /gene="PM1326"
     CDS             94. .1605
                     /gene="PM1326"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK03410.1"
                     /db_xref="GI:12721692"
                     /translation="MWIDVMTQRDCILKLKNIGKSFFDVPVLKNVDLDIYAGEVHCLV
                     GENGAGKSTLCKIIAGIYPFEQGEMFYEGKSYFPKTVKEAQSLGIGFIHQELMLVPQL
                     SVLENIFLGNEKTVFCGKMDWQTMRHKTQQIISELELDIDPDAKVADLSIAQQQMVEI
                     AKAVFSEYKVIIFDEPTSSISRKNTQTLFKIIHQLKEKNVAMIYISHRLDEFQYIADQ
                     VTVLRDGIRTGTMAYRDTSPDEIVKLMVGREINFAEYHRHYLPTREVLRVESLTNKHI
```

```
                KSSSFSLYEGEILGFAGLVGAGRTELLRAIFGADEASGAIYLNAQRVEIRSPEDAVKQ
                KIGFITEDRKQQGLILGLSIRENITLPILKRFWNGLFLNKKQERLVAEENRQKLRIVS
                HSQDQISQTLSGGNQQKVIIARWLESGVKILFFDEPTRGIDVGAKSEIYDLMRQFTDN
                GGAIVMVSSDLPELITMSDRVIVMRQGGIVREVKDKAEMTEENLMRLMIGVDE"
gene            1598. .2641
                /gene="rbsC_3"
                /note="PM1327"
CDS             1598. .2641
                /gene="rbsC_3"
                /codon_start=1
                /transl_table=11
                /product="RbsC"
                /protein_id="AAK03411.1"
                /db_xref="GI:12721693"
                /translation="MNRYCMTNLRKIFSKLGIGIILLLMIIGMSLSSDVFLSTNNIIN
                ILLQVSIICVISVGMTYVILTGGIDLSVGSIVALSAVCLGVFTHWGMDWLGENPSSFS
                VLMIVISAIIATIFVGILCGYVNGLVIVYGKVTPFITTLGMMGIARGLALTISDGKTI
                YNFPDTLRFLGNGRIALTETFALPVPVIIALLVVLVSFYVLTQTMFGRQIYALGGNRE
                AVRLSGINIEKLEIKAYVINGALAAIGAIILVGRLNAAQPIAGNGYELDAIAATVIGG
                TSLMGGVGSVVSTSIGALIMGVLQNGLTLLNVTSYLQRLIIGLVIILAVFLDQLRRGE
                VSTRRLKRLFFRE"
gene            3187. .4527
                /gene="PM1328"
CDS             3187. .4527
                /gene="PM1328"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03412.1"
                /db_xref="GI:12721694"
                /translation="MITIKKGLNLPISGSPEQVIRDGNAITEVALLGEEYVGMRPSMK
                VREGDVVKKGQVLFEDKKNPGVVFTAPASGTVTAIHRGAKRVLQSVVIKIEGNEQITF
                EKYTTEQLNQLTSEQVRQNLQASGLWTALRTRPFSKVPAVDATPVSIFVNAMDTNPLC
                ADPQVIVQQSAQAFEAGLTVLSRLHEGKVYLCKAANASIPSPSIANLDVKEFAGPHPA
                GLSGTHIHFIDPVSATKFVWYINYQDVIAVGKLFTTGELDVSRVVSLAGPQVKNPRLV
                RTVLGANLSQLTANEVKDGENRVISGSVLSGAKAAGPVDYLGRYALQVSVLEEGREKE
                FLGWIMPGANKYSLSRTVLGHFSKKLFNFTTALNGGERAMVPIGAYERVMPLDIIPTL
                LLRDLAAGDTDSAQALGCLELDEEDLALCTFVCPGKNEYGPLLRQALDKIEKEG"
gene            4530. .5762
                /gene="nqrB"
                /note="PM1329"
CDS             4530. .5762
                /gene="nqrB"
                /codon_start=1
                /transl_table=11
                /product="NqrB"
                /protein_id="AAK03413.1"
                /db_xref="GI:12721695"
                /translation="MGLKHFIEKIEPAFLPGGKYEKWYALFEATATFLYTPGTVTHKA
                SHVRDALDSKRMMVLVWLSLFPAMFYGMYNVGAQAILATDALGTLQQAISGHWQYALS
                HALGADLTLSAGWASKMLLGATYFLPIYLTIFLVGGFWEVLFAMVRKHEINEGFFVTS
                ILLALIVPPTLPLWQAALATTFGVVVAKEIFGGVGKNFMNPALAARAFLFFAYPAQIS
                GDTVWVAADGFSGATALSQWAVGGEAGLKHVVTGQPITWMDAFIGNIPGSIGEVSTLA
                LIIGAAIIVFARIASWRIIAGVMIGMIATSALFNLIGSSTNPLFAMPWYWHFVLGGFA
                LGMFFMATDPVSAAFTNKGKWWYGILIGAMAVLIRVVNPAYPEGMMLAILFANLFAPV
                FDYLVVQANIKRRRARNV"
gene            5755. .6537
                /gene="PM1330"
CDS             5755. .6537
                /gene="PM1330"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03414.1"
                /db_xref="GI:12721696"
                /translation="MFKNKDSVGGTLLVIILLSLACSIIVAGSAVLLKPTQIEQKELD
                KQKNILSVAGLLQPTTKNSEIKTIYANNIEARLVDLNTGDFAPAQPGFDAAKAVKNPA
                ESTALSAEDDVAGIRVRANFAEVYLVKNDAGQVTQVVLPFYGKGLWSMMYGFMSVQPD
                GNTVNGITYYDQGETPGLGGEIENPKWQAQFPGKKLYTADNQVGLYVGKGASANAEHG
                IDAISGSTLTSNGVNNSFKFWLGQKGFGPFLAKLKAGVLNNG"
gene            6530. .7156
                /gene="PM1331"
CDS             6530. .7156
                /gene="PM1331"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03415.1"
                /db_xref="GI:12721697"
                /translation="MADTKKLKGLLLSPVMDNNPIALQILGICSALAVTTKLETAIVM
                TFAVIFVTAFSNLFISLIRNYIPNSVRIIVQMAIIASLVILVDQILRAYAYGLSKQLS
                VFVGLIITNCIVMGRAEAFAMKSQPLESFVDGIGNGLGYGAILVSVAFIRELIGSGKL
                FGMTVFQTIQDGGWYQTNGLFLLAPSAFFIIGFIIWGIRTLKPNQVEK"
gene            7160. .7756
                /gene="nqrE"
                /note="PM1332"
CDS             7160. .7756
                /gene="nqrE"
                /codon_start=1
                /transl_table=11
                /product="NqrE"
                /protein_id="AAK03416.1"
                /db_xref="GI:12721698"
                /translation="MEHYISLFVKSVFIENMALSFFLGMCTFLAVSKKVSTAFGLGIA
                VIVVLGIAVPVNQLVYSFILKDSALVQGIDLSFLNFITFIGVIAALVQILEMVLDKYF
                PALYNALGIFLPLITVNCAIFGGVSFMVQRDYTFVESVVYGIGAGTGWMLAIVALAGI
                TEKMKYADVPAGLRGLGITFITVGLMALGFMSFSGIQL"
gene            7772. .8995
                /gene="nqr6"
                /note="PM1333"
CDS             7772. .8995
                /gene="nqr6"
                /codon_start=1
                /transl_table=11
                /product="Nqr6"
                /protein_id="AAK03417.1"
                /db_xref="GI:12721699"
                /translation="MEITLGIAMFTVIVLALAVIILFAKSKLVNSGDITIEINDDPSK
                AIHLPAGGKLLGALASQGIFVSSACGGGGSCGQCIVKVTEGGGDILPTELSHISKREA
                KEGYRLSCQVNVKNSMKVELPEEVFGVKKWECTVISNDNKATFIKELKLAIPEGEEVP
                FRAGGYIQIEAEPHTVAYKDFDIPEEYHEDWDKYNLWRYVSKVDEHIIRAYSMASYPE
                EKGIIMLNVRIATPPPNNPDAPPGQMSSYIWSLKAGDKVTISGPFGEFFAKETDAEMV
                FVGGGAGMAPMRSHIFDQLKRLKSKRKMSFWYGARSKREMFYVEDFDTLQAENDNFVW
                HVALSDPQPGDNWDGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPIMNASVIKMLKD
                LGVEDENILLDDFGG"
gene            9164. .10204
                /gene="apbE"
                /note="PM1334"
CDS             9164. .10204
                /gene="apbE"
                /codon_start=1
                /transl_table=11
                /product="ApbE"
                /protein_id="AAK03418.1"
                /db_xref="GI:12721700"
                /translation="MKHILNSIIIGFCSLCLLACHKSPEVVTITGKTMGTTYSVKYID
                DGQIKIKPIEMHGQIEEILKEVNNKMSTYLSTSELSQFNQNTQIDTPVEISADLAFVL
                KEAIRLHGVTEGALDVTVGPIVNLWGFGPEKRQPNETLEEQVEKRRAWVGIEKMKLTE
                QGDKFYLAKTVPQLYIDLSSIAKGFGVDKVAAHLATKGIANYLVEIGGEINAKGKNIE
                GKDWQIAIEKPNFDGSRAVSQIIGLKDFSMATSGNYRNYFEENGRRFSHEIDPKTGYP
                IEHRLASITVLADSTMTADGLSTGLYVLGEEKALEVAEKYNLLVYLISKTEKGFEAKM
                SSAFEKLLNEQH"
gene            10214. .10468
                /gene="PM1335"
CDS             10214. .10468
                /gene="PM1335"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03419.1"
                /db_xref="GI:12721701"
                /translation="MQLFLITFGIFLLIIAGMAIGYIIQRKTIAGSCGGISALGLKKV
                CDCEEPCDNLKAKLEAGDEEAQQEYNAKFAKNEPHFYEVK"
BASE COUNT      3052 a   1783 c   2451 g   3277 t
ORIGIN
```

```
  Query Match             41.7%;  Score 560;  DB 1;  Length 10563;
  Best Local Similarity   64.1%;  Pred. No. 1.3e-126;
  Matches  861;  Conservative    0;  Mismatches  480;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
Db    3187 ATGATTACGATTAAAAAAGGCTTGAATCTTCCTATTTCGGGAAGCCCAGAGCAGGTAATC 3246

Qy      61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
Db    3247 CGTGATGGCAATGCTATCACTGAAGTTGCTTTGCTTGGTGAAGAGTATGTGGGAATGCGC 3306

Qy     121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
Db    3307 CCTTCAATGAAGGTGCGTGAAGGTGATGTTGTGAAAAAAGGCCAAGTGCTTTTTGAAGAT 3366

Qy     181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
Db    3367 AAAAAGAACCCTGGTGTGGTTTTCACAGCCCCTGCAAGTGGTACTGTAACTGCGATTCAT 3426

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
Db    3427 CGTGGTGCTAAGCGAGTTTTACAATCTGTAGTAATTAAAATTGAAGGTAATGAACAAATT 3486

Qy     301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgccgccgc 360
Db    3487 ACTTTTGAAAAATATACAACAGAACAATTAAACCAACTTACTTCAGAGCAAGTTCGTCAA 3546

Qy     361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
Db    3547 AATTTACAAGCTTCTGGATTATGGACCGCACTTCGTACGCGTCCGTTCAGTAAAGTGCCT 3606

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
Db    3607 GCCGTCGATGCCACACCCGTATCGATTTTCGTGAATGCGATGGATACCAATCCATTATGT 3666

Qy     481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
Db    3667 GCCGATCCACAAGTGATCGTGCAACAATCTGCGCAAGCTTTTGAAGCGGGTCTGACTGTA 3726

Qy     541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
Db    3727 TTAAGTCGTTTACATGAGGGGAAAGTGTACCTGTGTAAAGCGGCGAATGCGAGCATTCCA 3786

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
Db    3787 AGCCCAAGCATTGCCAATCTTGACGTGAAAGAATTTGCAGGTCCACATCCAGCTGGGTTA 3846

Qy     661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
Db    3847 AGTGGTACGCATATTCACTTTATTGATCCTGTGAGTGCAACCAAATTTGTTTGGTATATC 3906

Qy     721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
Db    3907 AATTACCAAGATGTGATAGCGGTGGGTAAATTATTTACCACAGGTGAATTAGATGTTTCA 3966

Qy     781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
Db    3967 CGCGTAGTGTCCTTAGCTGGTCCACAAGTGAAAAATCCAAGATTAGTACGTACCGTATTA 4026

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
Db    4027 GGTGCAAACCTTTCCCAATTAACAGCGAATGAAGTGAAAGACGGTGAAAATCGCGTGATT 4086

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
Db    4087 TCAGGTTCAGTATTAAGTGGTGCAAAAGCTGCAGGTCCTGTTGATTATTTAGGTCGTTAC 4146

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db    4147 GCATTGCAAGTTTCAGTACTTGAAGAAGGTCGTGAGAAAGAATTTCTAGGTTGGATCATG 4206

Qy    1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
Db    4207 CCGGGTGCGAATAAATACTCACTTTCACGCACGGTGTTAGGGCATTTCTCTAAA---AAA 4263

Qy    1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
Db    4264 TTATTTAATTTTACCACCGCACTTAATGGCGGTGAGCGTGCGATGGTACCTATCGGTGCT 4323

Qy    1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
Db    4324 TATGAGCGTGTGATGCCATTGGATATTATTCCAACATTATTACTGCGTGATTTGGCAGCA 4383

Qy    1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
Db    4384 GGCGATACCGATTCTGCACAGGCGTTAGGGTGTTTAGAGCTAGATGAAGAAGATTTAGCC 4443

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
Db    4444 CTTTGTACCTTTGTTTGCCCTGGTAAAAATGAATATGGTCCGCTCTTACGTCAAGCTTTA 4503

Qy    1321 gaaaccattgagaaggaaggctga 1344
Db    4504 GATAAGATCGAGAAGGAAGGTTAA 4527


RESULT   8
APU24492
LOCUS       APU24492                1558 bp    DNA     linear   BCT 23-FEB-1996
DEFINITION  Actinobacillus pleuropneumoniae 48 kDa outer membrane protein
            (aopA) gene, complete cds.
ACCESSION   U24492
VERSION     U24492.1  GI:1185394
KEYWORDS    .
SOURCE      Actinobacillus pleuropneumoniae.
  ORGANISM  Actinobacillus pleuropneumoniae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Actinobacillus.
REFERENCE   1  (bases 1 to 1558)
  AUTHORS   Cruz,W.T., Nedialkov,Y.A., Thacker,B.J. and Mulks,M.H.
  TITLE     Molecular characterization of a common 48-kilodalton outer membrane
            protein of Actinobacillus pleuropneumoniae
  JOURNAL   Infect. Immun. 64 (1), 83-90 (1996)
  MEDLINE   96110919
REFERENCE   2  (bases 1 to 1558)
  AUTHORS   Mulks,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-1995) Martha H. Mulks, Microbiology, Michigan
            State University, 367 Giltner Hall, East Lansing, MI 48824, USA
FEATURES             Location/Qualifiers
     source          1. .1558
                     /organism="Actinobacillus pleuropneumoniae"
                     /strain="ISU158"
                     /serotype="5"
                     /db_xref="taxon:715"
     gene            95. .1444
                     /gene="aopA"
     CDS             95. .1444
                     /gene="aopA"
                     /note="similar to ngrA gene of Vibrio alginolyticus
                     encoding the alpha subunit of a sodium-translocating
                     NADH-uiquinone oxidoreductase, Genbank Accession Number
                     Z37111"
                     /codon_start=1
                     /transl_table=11
                     /product="48 kDa outer membrane protein"
                     /protein_id="AAC43631.1"
                     /db_xref="GI:1185395"
                     /translation="MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMK
                     VREGDVVKKGQVLFEDKKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITF
                     TRYEAAQLASLSAEQVKQNLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLA
                     ADPEVVLKEYETDFKDGLTVLTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGV
                     HPAGLVGTHIHFVDPVGATKQVWHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNP
                     RLVRTRLGANLSQLTANELNAGENRVISGSVLSGATAAGPVDYLGRYALQVSVLAEGR
```

```
                    EKELFGWIMPGSDKFSITRTVLGHFGKKLFNFTTAVHGGERAMVPIGAYERVMPLDII
                    PTLLLRDLAAGDTDSAQNLGCLELDEEDLALCTYVCPGKNNYGPMLRAALEKIEKEG"
     terminator     1462. .1499
BASE COUNT      432 a    299 c    375 g    452 t
ORIGIN


  Query Match           41.4%;  Score 556.6;  DB 1;  Length 1558;
  Best Local Similarity 64.8%;  Pred. No. 8.3e-126;
  Matches  877;  Conservative    0;  Mismatches  464;  Indels   12;  Gaps    3;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          ||||||| |||||| |||||  |  |  | || ||||| || | |||||  ||||   |
Db     95 ATGATTACAATCAAGAAAGGCTTGGATTTACCTATCGCAGGTACACCGGCACAAGTAATC 154

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           |  |||||    || |||  ||||| ||| ||||||||||||||||||  ||||||||
Db    155 CATAACGGCAATACCGTTAATGAAGTTGCGATGCTTGGCGAAGAATATGTGGGTATGCGT 214

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          || || ||||||||    ||||||||||  || |||||||| || || || ||||||||
Db    215 CCTTCAATGAAAGTTCGTGAAGGCGATGTAGTGAAAAAAGGTCAGGTTCTTTTTGAAGAT 274

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          |||||||||||||||||| || |||||||| || ||    || |   | |    | ||| ||
Db    275 AAAAAGAATCCGGGCGTTGTATTTACTGCTCCTGCAAGCGGTACCGTGGTTACTATTAAC 334

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          || || |||||||| || |||||||| |||||||||   ||||| ||  | ||  ||||
Db    335 CGCGGTGAAAAGCGTGTTCTTCAGTCGGTCGTGATTAAAGTTGAGGGTGATGAGCAAATT 394

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
             |||   |||||  | | |||       | ||      ||    | |||  |||||   |
Db    395 ACCTTTACTCGCTATGAAGCTGCGCAATTAGCGTCGCTATCAGCCGAGCAAGTGAAACAA 454

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
          || || ||| |||| ||||| |||||||||  | || || ||||||||||| ||  ||||
Db    455 AATCTTATCGAATCGGGTTTATGGACTGCATTCCGTACTCGTCCGTTCAGTAAGGTTCCC 514

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||  | || ||    |||   | ||||| |||||||| ||||| |||||||||  | ||
Db    515 GCATTGGACGCGATTCCGTCATCAATCTTTGTCAATGCAATGGATACCAATCCGTTAGCG 574

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          || || ||    ||  || | |||||   ||     ||||| ||| |||  |  . ||
Db    575 GCGGATCCGGAAGTGGTTTTAAAAGAGTACGAAACCGATTTTAAAGACGGTTTAACTGTT 634

Qy    541 ttgagccgtttgaccgaacgcaaaatccatgtt------tgtaaggcagctggcgcagac 594
          || | ||||||     |  |  ||| |  |||      ||||| |  || | |    |
Db    635 TTAACCCGTTTATTTAACGGTCAAAAACCGGTTTACTTATGTAAAGACGCGGACAGCAAT 694

Qy    595 gtgccgtctgaaaatgc---tgccaacatcgaaacacatgaattcggcggcccgcatcct 651
          | ||||        ||   ||     |||  ||   |    ||  ||||    ||||||
Db    695 ATCCCGTTAAGTCCGGCGATTGAAGGTATCACAATTAAGTCGTTTAGCGGTGTTCATCCT 754

Qy    652 gccggtttgagtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtg 711
          ||||||||    || ||||| || || ||  | || || || || ||||  |||   ||
Db    755 GCCGGTTTAGTCGGTACGCATATCCACTTTGTTGATCCCGTAGGGGCGACCAAACAAGTT 814

Qy    712 tggaccatcaattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctg 771
          |||  | | ||||||||||||||| ||| | || ||    || ||| | ||||||   ||
Db    815 TGGCACTTAAATTATCAAGATGTGATTGCTATCGGTAAATTATTTACTACAGGCGAACTC 874

Qy    772 aacaccgagcgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgt 831
             ||||| ||  | ||| | || |  ||| | ||||| || || || ||  |  | |||
Db    875 TTTACCGACCGTATTATTTCGCTTGCCGGTCCGCAAGTGAAAAATCCTCGTTTAGTACGT 934

Qy    832 accgttttgggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaac 891
          ||   | | ||||||||  | || ||| | || ||    || ||   || |   || |||
Db    935 ACGCGTCTTGGTGCGAATCTTTCCCAATTAACCGCAAATGAGTTAAATGCCGGTGAAAAC 994

Qy    892 cgcgtgatttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttg 951
          || |||||||| |||||||||||||| ||| |||| | |    ||  ||   ||||| ||

Db    995 CGTGTGATTTCAGGTTCGGTATTGAGCGGTGCGACTGCGGCGGGTCCGGTTGATTACTTA 1054

Qy    952 ggacgctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggc 1011
          || || |||       ||  | |||||  | |  |||||  ||    ||||||  | |||||
Db   1055 GGCCGTTACGCATTACAAGTATCCGTACTGGCGGAAGGTCGTGAGAAAGAGTTATTCGGT 1114

Qy   1012 tgggttgcgccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctg 1071
          ||| ||  ||||    |||| |||| |||||| || |||||     | |||||  |||
Db   1115 TGGATTATGCCGGGTTCGGATAAATTCTCCATTACCCGTACCGTGTTAGGCCACTTCGGT 1174

Qy   1072 aaaaacaaactcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccg 1131
          |||   ||| | || || ||||  || || ||  ||||||| || || || ||||||||
Db   1175 AAA---AAATTATTTAACTTCACTACGGCAGTTCACGGCGGTGAACGTGCAATGGTGCCA 1231

Qy   1132 attggtacttacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgat 1191
          || ||| | || ||||| || ||||| || |||||  | || ||  |  |  | || |||
Db   1232 ATCGGTGCGTATGAGCGTGTAATGCCGTTAGATATTATTCCTACGTTATTACTTCGTGAT 1291

Qy   1192 ttaatcgtcggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaa 1251
          |||   | ||||||||||||    || ||    || ||||| || ||  |||| ||||||
Db   1292 TTAGCTGCCGGCGATACCGATTCTGCACAAAACTTAGGTTGTTTAGAGCTGGATGAAGAA 1351

Qy   1252 gacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgc 1311
          ||  | || || || |  |  || |||||||| ||| |  | ||||| ||| |||| ||
Db   1352 GATTTGGCATTATGTACTTATGTTTGCCCGGGTAAAAATAACTACGGTCCGATGTTACGT 1411

Qy   1312 aaagtgctggaaaccattgagaaggaaggctga 1344
             | | | || |  || ||||||||||| | |
Db   1412 GCTGCGTTAGAGAAGATCGAGAAGGAAGGTTAA 1444


RESULT   9
U32702
LOCUS       U32702                 13299 bp    DNA     linear   BCT 29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 17 of 163 of the complete genome.
ACCESSION   U32702 L42023
VERSION     U32702.1  GI:1573118
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
  ORGANISM  Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 13299)
  AUTHORS   Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
REFERENCE   2  (bases 1 to 13299)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
REFERENCE   3  (bases 1 to 13299)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 13299)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
```

```
  REMARK     The H. influenzae sequence has been updated by R. Fleischmann. New
             database matches have been assigned, product names have been
             improved, and a number of frame shifts have been corrected. We
             gratefully acknowledge the work of Tatusov et. al. We have
             incorporated their annotation into the /notes fields of the
             corresponding H. influenzae genes
REFERENCE    5  (bases 1 to 13299)
  AUTHORS    White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
             Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
             Medical Center Dr, Rockville, MD 20850, USA
  REMARK     The whole genome was shifted by 588 nucleotides for a new start
COMMENT      On Sep 30, 1996 this sequence version replaced gi:1220812.
FEATURES             Location/Qualifiers
     source          1. .13299
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            complement(85. .1455)
                     /gene="HI0161"
     CDS             complement(85. .1455)
                     /gene="HI0161"
                     /note="similar to SP:P06715 GB:M13141 PID:146248
                     PID:466637 GB:U00096 percent identity: 74.17; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glutathione reductase (gor)"
                     /protein_id="AAC21833.1"
                     /db_xref="GI:1573119"
                     /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                     VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                     LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                     DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                     LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                     LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                     KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                     QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                     EFVTMR"
     gene            complement(1671. .2270)
                     /gene="HI0162"
     CDS             complement(1671. .2270)
                     /gene="HI0162"
                     /note="similar to GB:U00096 PID:1773118 PID:1786637
                     percent identity: 25.82; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="lipoprotein, putative"
                     /protein_id="AAC21834.1"
                     /db_xref="GI:1573120"
                     /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                     VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                     QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                     VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
     gene            2364. .2675
                     /gene="HI0163"
     CDS             2364. .2675
                     /gene="HI0163"
                     /note="similar to PID:663269 percent identity: 46.53;
                     identified by sequence similarity; putative".
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, putative (bolA)"
                     /protein_id="AAC21835.1"
                     /db_xref="GI:1573121"
                     /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                     SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                     R"
     gene            2966. .4309
                     /gene="HI0164"
     CDS             2966. .4309
                     /gene="HI0164"
                     /note="similar to PID:663269 PID:663270 percent identity:
                     64.96; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreducatase subunit A, Na
                     translocating"
                     /protein_id="AAC21836.1"
                     /db_xref="GI:1573122"
                     /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                     VREGDVVKKGQVLFEDKKNPGVIFTAPASGTITAINRGEKRVLQSVVINVEGDEKITF
                     AKYSTEQLNTLSSEQVKQNLIESGLWTALRTRPFSKVPSIESEASSIFVNAMDTNPLA
                     ADPSVVLKEYSQDFTNGLTVLSRLFPSKPLHLCKAGDSNIPTADLENLQIHDFTGVHP
                     AGLVGTHIHFIDPVGIQKTVWHINYQDVIAVGKLFTTGELYSERVISLAGPQVKEPRL
                     VRTTIGANLSQLTQNELSAGKNRVISGSVLCGQIAKDSHDYLGRYALQVSVIAEGNEK
                     EFFGWIMPQANKYSVTRTVLGHFSKKLFNFTTSENGGERAMVPIGSYERVMPLDILPT
                     LLLRDLIVGDTDGAQELGCLELDEEDLALCSFVCPGKYEYGSILRQVLDKIEKEG"
     gene            4312. .5547
                     /gene="HI0166"
     CDS             4312. .5547
                     /gene="HI0166"
                     /note="similar to PID:663269 PID:663271 percent identity:
                     71.33; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, subunit B
                     (nqrB)"
                     /protein_id="AAC21837.1"
                     /db_xref="GI:1573123"
                     /translation="MGLKNLFEKMEPAFLPGGKYSKLYPIFESIYTLLYTPGTVTHKN
                     THVRDALDSKRMMITVFLALFPAIFYGMYNVGNQAIPALNQLGNLDQLIANDWHYALA
                     SSLGLDLTANATWGSKMALGAIFFLPIYLVVFTVCTIWELLFSVVRGHEVNEGMFVST
                     ILFALIVPPTLPLWQAALGITFGIIVAKEIFGGVGRNFMNPALAGRAFLFFAYPAQIS
                     GDTVWTAADGFSGATALSQWSQGGQGALQHTVTGAPITWMDAFVGNLPGSMGEVSTLA
                     ILIGGAVIVFTRIAAWRIIAGVMIGMIATSTLFNLIGSETNPMFSMPWHWHFVLGGFA
                     LGMVFMATDPVSASFTNTGKWWYGALIGVMAVLIRTVNPAYPEGMMLAILFANLFAPI
                     FDYIVVQANIKRRRARTNG"
     gene            5540. .6274
                     /gene="HI0167"
     CDS             5540. .6274
                     /gene="HI0167"
                     /note="similar to PID:663269 PID:663272 PID:893413 percent
                     identity: 54.36; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating"
                     /protein_id="AAC21838.1"
                     /db_xref="GI:1573124"
                     /translation="MAKFNKDSVGGTILVVLLLSLVCSIIVAGSAVMLKPAQEEQKLL
                     DKQKNILNVAGLLQENTNVKETYAKFIEPRFVDLATGEYTQQADDSQQAIPADADKAR
                     IRSRSKTTEVYLVKDEQGQTQQVILPIYGTGLWSVMYGLVSVQPDGNTINGITYYQHG
                     ETPGLGGEIENPNWASLFKGKKLFDEQHQPAIRIVKGQAPQDEHSIDGLSGATLTGNG
                     VQGTFNYWFSKDGFGPYLEKLHSGAN"
     gene            6274. .6900
                     /gene="HI0168"
     CDS             6274. .6900
                     /gene="HI0168"
                     /note="similar to PID:663269 PID:663273 PID:893414 percent
                     identity: 74.88; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating"
                     /protein_id="AAC21839.1"
                     /db_xref="GI:1573125"
                     /translation="MSGKTSYKDLLLAPIAKNNPIALQILGICSALAVTTKLETAFVM
                     AIAVTLVTGLSNLFVSLIRNYIPNSIRIIVQLAIIASLVIVVDQILKAYAYGLSKQLS
                     VFVGLIITNCIVMGRAEAFAMKSPPVESFVDGIGNGLGYGSMLIIVAFFRELIGSGKL
                     FGMTIFETIQNGGWYQANGLFLLAPSAFFIIGFVIWGLRTWKPEQQEK"
     gene            6904. .7500
                     /gene="HI0170"
```

```
    CDS             6904. .7500
                    /gene="HI0170"
                    /note="similar to PID:893416 PID:893415 percent identity:
                    84.34; identified by sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="NADH:ubiquinone oxidoreductase, Na
                    translocating"
                    /protein_id="AAC21840.1"
                    /db_xref="GI:1573126"
                    /translation="MEHYISLFVKAVFIENMALSFFLGMCTFLAVSKKVSPAFGLGIA
                    VTFVLGIAVPVNQLIYANVLKENALIEGVDLSFLNFITFIGVIAGLVQILEMVLDKFM
                    PSLYNALGIFLPLIAVNCAIFGGVSFMVQRDYNFPESIVYGFGSGLGWMLAIVALAGL
                    TEKMKYADIPAGLKGLGITFISVGLMALGFMSFSGIQL"

  Query Match             39.0%;  Score 524.6;  DB 1;  Length 13299;
  Best Local Similarity   62.9%;  Pred. No. 6.5e-118;
  Matches  847;  Conservative    0;  Mismatches  494;  Indels    6;  Gaps    2;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          ||||||| ||| || |||||| |  | || || || ||||| | ||| |  ||||   |
Db   2966 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 3025

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            |   ||||   ||  | |   |||| ||| | || || || || ||||  || |||||
Db   3026 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 3085

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          || || ||||| ||    |||||||||  || |||||||| ||||| || |||||||||
Db   3086 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTGAAGAC 3145

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          ||||| ||||| || ||  | ||||| || || ||    || |  ||| | || ||  |
Db   3146 AAAAAAAATCCTGGTGTAATTTTTACAGCCCCTGCAAGCGGTACCATCACTGCAATCAAT 3205

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          |||||||||||| || ||| | || || || ||| || ||||| | ||  ||||||
Db   3206 CGTGGCGAAAAACGTGTATTACAATCTGTTGTTATTAATGTAGAAGGTGATGAAAAAATC 3265

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
            |||| |   ||     | |||    ||   |   |      ||| ||||    |
Db   3266 ACTTTTGCAAAATATAGCACAGAACAATTGAATACACTTTCTTCTGAACAAGTAAAACAA 3325

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
          ||| ||||  |||| || || ||||| ||  | || || ||||| || || ||| ||||
Db   3326 AACTTGATAGAATCTGGCTTATGGACGGCATTACGTACTCGTCCATTTAGTAAAGTTCCA 3385

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
            || | ||    ||  | |   | || || || ||||||||||| ||||||||  | ||
Db   3386 TCCATTGAAAGTGAAGCATCTTCTATTTTTGTAAATGCGATGGATACCAATCCATTAGCA 3445

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          ||||| ||  | ||| |  | ||||||    |  | || ||||    ||| |    |||
Db   3446 GCCGATCCCTCTGTCGTATTAAAAGAATATTCGCAAGACTTCACTAACGGTTTAACCGTA 3505

Qy    541 ttgagccgt---ttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtg 597
          || || |||   ||  |   |     | | ||  | ||||| |||| |   |  | |  |
Db   3506 TTAAGTCGTCTATTCCCTTCAAAACCATTACACTTATGTAAAGCAGGCGATTCTAATATT 3565

Qy    598 ccgtctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggt 657
          ||  |||   ||  ||  ||  |  | |  ||||| ||    ||    ||||| || |||
Db   3566 CCAACTGCTGATCTTGAAAATTTACAGATTCATGATTTTACTGGTGTTCATCCAGCTGGT 3625

Qy    658 ttgagtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggacc 717
            |   || ||||| ||||| || || || || ||||     | |||||| || |||
Db   3626 CTTGTAGGGACGCATATTCACTTTATTGATCCTGTCGGTATTCAAAAAACTGTATGGCAT 3685

Qy    718 atcaattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacacc 777
          ||||||||||||||||| ||  |  | ||    || ||| ||||||||   ||  |  |
Db   3686 ATCAATTATCAAGATGTGATCGCTGTAGGTAAATTATTTACAACAGGCGAACTTTATTCA 3745

Qy    778 gagcgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgtt 837
          || || || ||| | || |  ||| ||||||| ||  |||| ||  |  | |||||

Db   3746 GAACGAGTAATTTCTCTTGCAGGTCCTCAAGTGAAAGAACCTCGTTTAGTACGTACAACA 3805

Qy    838 ttgggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtg 897
            | ||||| ||  |  || ||| | ||       |||||   ||      | || || ||
Db   3806 ATCGGTGCTAATCTTTCTCAATTAACCCAAAATGAATTAAGTGCAGGTAAAAATCGTGTA 3865

Qy    898 atttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgc 957
          ||||| || || ||  |    ||    ||  || |||   | || |||||||| || |||
Db   3866 ATTTCTGGCTCAGTTCTTTGTGGTCAAATAGCAAAAGATTCACATGATTATTTAGGTCGC 3925

Qy    958 taccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggtt 1017
          ||       ||  | || || ||||  |||||       |||||  | || || ||| ||
Db   3926 TATGCTTTGCAAGTATCAGTAATCGCTGAAGGTAATGAAAAAGAATTTTTTGGTTGGATT 3985

Qy   1018 gcgccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaac 1077
            ||| ||  |  | ||||| ||  | || |||||     | ||||| |||   |  ||
Db   3986 ATGCCACAAGCCAATAAATATTCTGTGACTCGTACCGTATTAGGCCACTTC---AGTAAA 4042

Qy   1078 aaactcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggt 1137
          ||| | || || || |  | |  || || ||||| || || || |||||||| || |||
Db   4043 AAATTGTTTAATTTTACTACTTCAGAAAATGGTGGCGAACGTGCAATGGTGCCAATCGGT 4102

Qy   1138 acttacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatc 1197
          |  || |||||||| ||||| ||||| ||  | || ||  |  | || || |||||||||
Db   4103 AGCTATGAGCGCGTAATGCCATTGGACATTTTACCTACATTATTATTACGTGATTTAATC 4162

Qy   1198 gtcggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctc 1257
            || ||||||||||  | || || |  || |||||  ||||||| ||||||||||| |
Db   4163 GTGGGCGATACCGATGGTGCTCAAGAGTTAGGTTGTCTGGAATTAGACGAAGAAGACTTG 4222

Qy   1258 gctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtg 1317
          || || ||    ||||| || || |||||||| ||||||||  |  | |||||  ||||
Db   4223 GCGTTATGTTCTTTCGTTTGTCCAGGCAAATATGAATACGGTTCAATCTTGCGTCAAGTC 4282

Qy   1318 ctggaaaccattgagaaggaaggctga 1344
            | || |  ||||||||||||||| | |
Db   4283 TTAGATAAGATTGAGAAGGAAGGTTAA 4309


RESULT  10
AF165980
LOCUS       AF165980               6048 bp    DNA     linear   BCT 13-NOV-2001
DEFINITION  Vibrio harveyi Na+-translocating NADH-quinone oxidoreductase
            complex operon, complete sequence.
ACCESSION   AF165980
VERSION     AF165980.1  GI:6531420
KEYWORDS    .
SOURCE      Vibrio harveyi.
  ORGANISM  Vibrio harveyi
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 6048)
  AUTHORS   Zhou,W., Bertsova,Y.V., Feng,B., Tsatsos,P., Verkhovskaya,M.L.,
            Gennis,R.B., Bogachev,A.V. and Barquera,B.
  TITLE     Sequencing and preliminary characterization of the
            Na+-translocating NADH: ubiquinone oxidoreductase from Vibrio
            harveyi
  JOURNAL   Biochemistry 38 (49), 16246-16252 (1999)
  MEDLINE   20056044
   PUBMED   10587447
REFERENCE   2  (bases 1 to 6048)
  AUTHORS   Zhou,W., Feng,B., Gennis,R.B. and Barquera,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1999) Biochemistry, University of Illinois, 600
            South Mathews Avenue, Urbana, IL 61801, USA
FEATURES             Location/Qualifiers
     source          1. .6048
                     /organism="Vibrio harveyi"
                     /strain="BB120"
                     /db_xref="taxon:669"
     gene            169. .1509
                     /gene="nqrA"
     gene            169. .6022
                     /gene="Na+-translocating NADH-quinone oxidoreductase
```

```
                complex operon"
     CDS        169. .1509
                /gene="nqrA"
                /codon_start=1
                /transl_table=11
                /product="NqrA"
                /protein_id="AAF15411.1"
                /db_xref="GI:6531421"
                /translation="MITIKKGLALPIAGAPSQVINDGKTIKKVALLGEEYVGMRPTMH
                VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEINRGAKRVLQSVVIEVAGEEQVTF
                DKFEAAQLAGLDREVIKTQLVESGLWTALRTRPFSKVPAIESATKAIFVTAMDTNPLA
                AQPELIITEQQEAFVAGLDILSALTEGKVYVCKSGTSLPSSSQSNVEEHVFDGPHPAG
                LAGTHMHFLYPVNAENVAWSINYQDVIAFGQLFLTGELYTDRVVSLAGPVVNNPRLLR
                TVVGASLEDLTDSELMPGEVRVISGSVLSGTQASGPHAYLGRYHQQVSVLREGRDKEL
                FGWATPGKNKFSITKSFLGHIFKGQLFNMTTTTNGSDRAMVPIGNYERVMPLDMEPTL
                LLRDLCAGDTDSAQTLGALELDEEDLALCTFVCPGKYEYGQLLRECLDTIVKEG"
     gene       1513. .2754
                /gene="nqrB"
     CDS        1513. .2754
                /gene="nqrB"
                /codon_start=1
                /transl_table=11
                /product="NqrB"
                /protein_id="AAF15412.1"
                /db_xref="GI:6531422"
                /translation="MLKKFIEDIEHHFEPGGKHEKWFALYEAAATLFYTPGLVTKRSS
                HVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGDQLATVIAGNWHYW
                LTEMLGGTIGAEAGVGSMMLLGATYFLPIYATVFIVGGFWEVLFCMVRKHEVNEGFFV
                TSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPALAGRAFLFFAYPAQ
                ISGDVVWTAADGFSGATALSQWAHRGSGALINNITGAPITWMDAFIGNIPGSIGEVST
                LALMIGAAMIVYMRIASWRIIAGVMIGMIAVSTLFNVIGSDTNPMFNMPWHWHLVLGG
                FAFGMFFMATDPVSASFTNSGKWWYGILIGAMCVMIRVVNPAYPEGMMLAILFANLFA
                PLFDHVVIEKNIKRRLARYGK"
     gene       2744. .3529
                /gene="nqrC"
     CDS        2744. .3529
                /gene="nqrC"
                /codon_start=1
                /transl_table=11
                /product="NqrC"
                /protein_id="AAF15413.1"
                /db_xref="GI:6531423"
                /translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDQQKANAVLD
                KQSKIVEVAGIDAEGKKVPELFAEYIEPRLVDFKTGDFVEKAEDGSTAANYDQRKAAK
                DPAESIKLTADEDKAKILRRANTGIVYLVKNGDDISKVIIPVHGNGLWSMMYAFVAVE
                TDGNTVSGITYYEQGETPGLGGEVENPVWRAQFVGKKLFDENHKPAIKIVKGGAPEGS
                EHGVDGLSGATLTGNGVQGTFDFWLGDMGFGPFLAKVRDGGLN"
     gene       3529. .4161
                /gene="nqrD"
     CDS        3529. .4161
                /gene="nqrD"
                /codon_start=1
                /transl_table=11
                /product="NqrD"
                /protein_id="AAF15414.1"
                /db_xref="GI:6531424"
                /translation="MSSAQNIKKSIMAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
                VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                FGMEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
     gene       4169. .4765
                /gene="nqrE"
     CDS        4169. .4765
                /gene="nqrE"
                /codon_start=1
                /transl_table=11
                /product="NqrE"
                /protein_id="AAF15415.1"
                /db_xref="GI:6531425"
                /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGVA
                VVVVLTIAVPVNNLVYNLVLKENALVEGVDLSFLNFITFIGVIAALVQILEMVLDRFF
                PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYNFAESIVYGFGSGVGWMLAIVALAGI
                REKMKYSDVPPGLRGLGITFITVGLMALGFMSFSGVQL"
     gene       4799. .6022
                /gene="nqrF"
     CDS        4799. .6022
                /gene="nqrF"
                /codon_start=1
                /transl_table=11
                /product="NqrF"
                /protein_id="AAF15416.1"
                /db_xref="GI:6531426"
                /translation="MDIILGVVMFTLIVLALVLVILFAKSKLVPTGDITISVNDDPSL
                AIVTQPGGKLLSALAGAGVFVSSACGGGGSCGQCRRKIKSGGGDILPTELDHITKGEA
                REGERLACQVAMKTDMDIELPEEIFGVKKWECTVISNDNKATFIKELKLQIPDGESVP
                FRAGGYIQIEAPAHHVKYADYDIPEEYREDWEKFNLFRYESKVNEETIRAYSMANYPE
                EHGIIMLNVRIATPPPNNPDVAPGIMSSFIWSLKEGDKCTISGPFGEFFAKDTDAEMV
                FVGGGAGMAPMRSHIFDQLKRLHSKRKMSFWYGARSKREMFYVEDFDGLAADNDNFVW
                HCALSDPLPEDNWDGYTGFIHNVLYENYLRDHEAPEDCEYYMCGPPMMNAAVIGMLKD
                LGVEDENILLDDFGG"
BASE COUNT     1571 a   1351 c   1455 g   1671 t
ORIGIN
```

```
  Query Match             37.6%;  Score 505.6;  DB 1;  Length 6048;
  Best Local Similarity   61.5%;  Pred. No. 2.9e-113;
  Matches  827;  Conservative    0;  Mismatches  514;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
           ||||||| ||| || || ||  |    || || ||||| ||    ||   || |  ||
Db     169 ATGATTACAATAAAGAAGGGCTTGGCTCTTCCTATCGCAGGAGCTCCATCCCAGGTGATT 228

Qy      61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
             | || ||   | |||| |   ||||||| |||||||||||||| || |  || |||||
Db     229 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 288

Qy     121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||  | ||| | |||   |  |||||||  || |||||||  ||||| || |||||||||
Db     289 CCTACCATGCATGTCCGCGTTGGCGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAC 348

Qy     181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
           |||||||| || |||||   ||| ||||| || ||  | || ||| |   ||  ||| ||
Db     349 AAAAAGAACCCTGGCGTTAAGTTCACTGCTCCAGCAGCCGGTAAAGTGATCGAAATTAAC 408

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           |||||||  ||||| |||||||| ||||| ||||||| || | |||| | ||  |  |
Db     409 CGTGGCGCTAAGCGTGTACTTCAATCAGTTGTGATTGAAGTAGCAGGCGAAGAGCAGGTT 468

Qy     301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
             ||||||   | || | | |    || |||    ||   ||||||  |
Db     469 ACGTTTGATAAGTTCGAAGCAGCTCAACTAGCAGGTCTAGATCGCGAAGTGATCAAGACT 528

Qy     361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
             |  || |  |||| ||  | ||||| ||  | || || ||||||||||||||  ||||
Db     529 CAATTGGTTGAATCTGGCCTTTGGACCGCTTTACGTACTCGTCCGTTCAGCAAGGTTCCA 588

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           ||  |||| | | |    || || ||||| | |||||||||| || ||||| || ||
Db     589 GCAATCGAGTCTGCTACTAAGGCTATTTTCGTAACTGCGATGGATACTAATCCACTAGCA 648

Qy     481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
           ||  | |||  | | || ||||  |||      || |  |||      || ||     |
Db     649 GCACAGCCTGAGTTGATCATCACTGAACAACAAGAAGCATTCGTTGCAGGTCTAGACATT 708

Qy     541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
             |    |||| ||| ||||  |  | || ||||||    ||||| |   | ||||
Db     709 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAAG---TCTGGCACTAGCTTGCCA 765

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
                 |    | ||| | |||  ||| |  ||||  || || || || || ||| |
Db     766 AGCTCTTCTCAGTCAAACGTAGAAGAACACGTCTTCGATGGTCCTCACCCAGCAGGTCTT 825

Qy     661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
             |||||| ||||| |||||| |  | || ||    ||  | ||    | ||||| ||||
Db     826 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCGTGGAGCATC 885
```

```
Qy    721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
          || || ||||| || ||  |  | || |   | ||   ||| ||    ||  |||| ||
Db    886 AACTACCAAGACGTTATCGCGTTCGGTCAGCTTTTCCTAACGGGTGAGCTTTACACTGAT 945

Qy    781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          || ||  |  | |||| |||| ||   || ||||| || |||||  | ||||| ||| |
Db    946 CGTGTTGTGTCTCTAGCTGGTCCTGTTGTGAACAACCCACGCCTTCTTCGTACAGTTGTA 1005

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
          |||||     |     |  | ||||   |||| ||| |        ||    || ||||||
Db   1006 GGTGCTTCTCTTGAAGACCTAACTGACAGCGAGTTGATGCCTGGTGAAGTTCGTGTGATT 1065

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          || || |||||| |    ||  |     |    ||  | || |  ||  | || || |||
Db   1066 TCAGGCTCGGTACTTTCTGGTACTCAAGCGTCTGGTCCACATGCCTACCTAGGTCGTTAC 1125

Qy    961 cacaatcagatttccgttatcgaagaagccgcagcaaagagctgttcggctgggttgcg 1020
          ||| | ||  |||| ||  |    ||||| ||   |||||| ||||| ||||||| | |
Db   1126 CACCAACAAGTTTCTGTGTTACGCGAAGGTCGTGACAAAGAACTGTTTGGCTGGGCTACA 1185

Qy   1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
          ||        |||| | ||| ||||||    |   ||| || ||  || | ||     |
Db   1186 CCTGGTAAAAACAAGTTCTCAATCACGAAATCTTTCCTTGGTCACATCTTCAAGGGTCAG 1245

Qy   1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
           | |||||  | |  ||| |   ||||||  | ||||| || ||||| || || || |
Db   1246 TTGTTCAACATGACAACAACGACCAACGGTAGTGACCGTGCAATGGTTCCAATCGGCAAC 1305

Qy   1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
          |||||||||||| |||||  | |||||    || ||  ||||  | || ||| ||   |
Db   1306 TACGAGCGCGTAATGCCGCTTGATATGGAACCGACTTTGCTACTTCGTGATCTATGTGCA 1365

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
          |||||||| |||| || ||  || |||||    | ||| |||| |||||||| || ||
Db   1366 GGCGATACAGACAGTGCTCAAACACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 1425

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          ||||| | ||| || || || || |||||||| ||||| | |||  | ||  ||   ||
Db   1426 TTGTGTACCTTTGTATGTCCTGGTAAATACGAGTACGGTCAGCTACTTCGTGAATGCCTA 1485

Qy   1321 gaaaccattgagaaggaaggctga 1344
          || ||||| | |||||||||| | |
Db   1486 GATACCATCGTGAAGGAAGGGTAA 1509
```

```
RESULT  11
AF117331
LOCUS       AF117331                5916 bp    DNA     linear   BCT 13-MAY-1999
DEFINITION  Vibrio cholerae N16961 Na+-translocating NADH-ubiquinone
            oxidoreductase enzyme complex, complete sequence.
ACCESSION   AF117331
VERSION     AF117331.1  GI:4808841
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 5916)
  AUTHORS   Hase,C.C. and Mekalanos,J.J.
  TITLE     Effects of changes in membrane sodium flux on virulence gene
            expression in Vibrio cholerae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3183-3187 (1999)
  MEDLINE   99179036
REFERENCE   2  (bases 1 to 5916)
  AUTHORS   Hase,C.C. and Mekalanos,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1998) Microbiology and Molecular Genetics,
            Harvard Medical School, 200 Longwood Ave., Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .5916
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /db_xref="taxon:666"
                     /note="biotype: El Tor"
     gene            34. .1374
                     /gene="nqrA"
     CDS             34. .1374
                     /gene="nqrA"
                     /note="alpha subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrA"
                     /protein_id="AAD29962.1"
                     /db_xref="GI:4808842"
                     /translation="MITIKKGLDLPIAGTPSQVISDGKAIKKVALLGEEYVGMRPTMH
                     VRVGDEVKKAQILFEDKKNPGVKFTSPVSGKVVEINRGAKRVLQSVVIEVAGDDQVTF
                     DKFEANQLASLNRDAIKTQLVESGLWTAFRTRPFSKVPAIDSTSEAIFVTAMDTNPLA
                     AEPTVVINEQSEAFVAGLDVLSALTTGKVYVCKKGTSLPRSQQPNVEEHVFDGPHPAG
                     LAGTHMHFLYPVSADHVAWSINYQDVIAVGQLFLTGELYTQRVVSLAGPVVNKPRLVR
                     TVMGASLEQLVDSEIMPGEVRIISGSVLSGTKATGPHAYLGRYHLQVSVLREGRDKEL
                     FGWAMPGKNKFSVTRSFLGHLFKGQVYNMTTTTNGSDRSMVPIGNYEKVMPLDMEPTL
                     LLRDLCAGDSDSAVRLGALELDEEDLALCTFVCPGKYEYGQLLRECLDKIEKEG"
     gene            1380. .2627
                     /gene="nqrB"
     CDS             1380. .2627
                     /gene="nqrB"
                     /note="a subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrB"
                     /protein_id="AAD29963.1"
                     /db_xref="GI:4808843"
                     /translation="MGLKKFLEDIEHHFEPGGKHEKWFALYEAAATLFYTPGLVTKRS
                     SHVRDSVDLKRIMIMVWLAVFPAMFWGMYNAGGQAIAALNHLYSGDQLAAIVAGNWHY
                     WLTEMLGGTMSSDAGWGSKMLLGATYFLPIYATVFIVGGFWEVLFCMVRKHEVNEGFF
                     VTSILFALIVPPTLPLWQAALGITFGVVVAKEVFGGTGRNFLNPALAGRAFLFFAYPA
                     QISGDLVWTAADGYSGATALSQWAQGGAGALINNATGQTITWMDAFIGNIPGSIGEVS
                     TLALMIGAAFIVYMGIASWRIIGGVMIGMILLSTLFNVIGSDTNAMFNMPWHWHLVLG
                     GFAFGMFFMATDPVSASFTNSGKWAYGILIGVMCVLIRVVNPAYPEGMMLAILFANLF
                     APLFDHVVVERNIKRRLARYGKQ"
     gene            2614. .3387
                     /gene="nqrC"
     CDS             2614. .3387
                     /gene="nqrC"
                     /note="gamma subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrC"
                     /protein_id="AAD29964.1"
                     /db_xref="GI:4808844"
                     /translation="MASNNDSIKKTLFVVIALSLVCSIIVSAAAVGLRDKQKENAALD
                     KQSKILQVAGIEAKGSKQIVELFNKSIEPRLVDFNTGDFVEGDAANYDQRKAAKEASE
                     SIKLTAEQDKAKIQRRANVGVVYLVKDGDKTSKVILPVHGNGLWSMMYAFVAVETDGN
                     TVSGLTYYEQGETPGLGGEVENPAWRAQWVGKKLFDENHKPAIKIVKGGAPQGSEHGV
                     DGLSGATLTSNGVQNTFDFWLGDMGFGPFLTKVRDGGLN"
     gene            3387. .4019
                     /gene="nqrD"
     CDS             3387. .4019
                     /gene="nqrD"
                     /note="b subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrD"
                     /protein_id="AAD29965.1"
                     /db_xref="GI:4808845"
                     /translation="MSSAKELKKSVLAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVMFVTALSNFFVSLIRNHIPNSVRIIVQMAIIASLVIVVDQILKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSEPIPSFIDGIGNGLGYGFVLMTVGFFRELLGSGKL
                     FGLEVLPLISNGGWYQPNGLMLLAPSAFFLIGFMIWAIRTFKPEQVEAKE"
     gene            4027. .4623
                     /gene="nqrE"
     CDS             4027. .4623
                     /gene="nqrE"
                     /note="c subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrE"
```

```
                     /protein_id="AAD29966.1"
                     /db_xref="GI:4808846"
                     /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGIA
                     VIVVLTISVPVNNLVYNLVLKPDALVEGVDLSFLNFITFIGVIAALVQILEMILDRFF
                     PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYSFAESVVYGFGSGVGWMLAIVALAGI
                     REKMKYSDVPPGLRGLGITFITAGLMALGFMSFSGVQL"
     gene            4657. .5883
                     /gene="nqrF"
     CDS             4657. .5883
                     /gene="nqrF"
                     /note="beta subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrF"
                     /protein_id="AAD29967.1"
                     /db_xref="GI:4808847"
                     /translation="MSTIIFGVVMFTLIILALVLVILFAKSKLVPTGDITISINGDPE
                     KAIVTQPGGKLLTALAGAGVFVSSACGGGGSCGQCRVKIKSGGGDILPTELDHISKGE
                     AREGERLACQVAVKADMDLELPEEIFGVKKWECTVISNDNKATFIKELKLAIPDGESV
                     PFRAGGYIQIEAPAHHVKYADFDVPEKYRGDWDKFNLFRYESKVDEPIIRAYSMANYP
                     EEFGIIMLNVRIATPPPNNPNVPPGQMSSYIWSLKAGDKCTISGPFGEFFAKDTDAEM
                     VFIGGGAGMAPMRSHIFDQLKRLKSKRKMSYWYGARSKREMFYVEDFDGLAAENDNFV
                     WHCALSDPQPEDNWTGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPMMNAAVINMLK
                     NLGVEEENILLDDFGG"
BASE COUNT     1441 a   1276 c   1550 g   1649 t
ORIGIN
```

```
  Query Match             36.4%;  Score 489.6;  DB 1;  Length 5916;
  Best Local Similarity   60.8%;  Pred. No. 2.4e-109;
  Matches  817;  Conservative    0;  Mismatches  524;  Indels   3;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          ||||||| ||| ||||| ||  |  |||| || ||||| || |  ||    || |   ||
Db     34 ATGATTACAATAAAAAAGGGATTGGACCTTCCTATCGCAGGAACTCCTTCCCAGGTGATT 93

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            || ||    ||||| |   ||||||| ||||||||||||||| || |  || |||||
Db     94 AGTGATGGTAAAGCCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 153

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||  | ||| | |||   | ||| ||||  || |||||||  ||  | || ||||||||
Db    154 CCTACTATGCATGTCCGCGTAGGTGATGAAGTAAAAAAAGCTCAGATTCTTTTTGAAGAT 213

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          || |||||||||||| |||   |||||| | || |     || ||| | |  |  ||  |
Db    214 AAGAAGAATCCGGGTGTGAAATTTACTTCACCAGTCAGCGGTAAAGTGGTTGAAATCAAT 273

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          |||||  | |||||||||| |||||||| || |||||||  || |  ||  |  ||| ||  |
Db    274 CGTGGTGCAAAGCGCGTTCTTCAGTCTGTAGTGATTGAAGTGGCGGGTGATGACCAAGTG 333

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
            || ||    |  | | | |    | || |  || |  |||| |   |  .
Db    334 ACCTTCGACAAGTTTGAAGCCAATCAATTAGCCAGTTTGAATCGCGATGCGATCAAAACT 393

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
            | ||| |  | || ||| | ||||||||  | || || ||||||||||||||  ||||
Db    394 CAGCTGGTTGAGTCTGGTCTTTGGACTGCTTTCCGTACTCGTCCGTTCAGCAAGGTTCCA 453

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||| | ||| ||    |    ||||| || |||| ||| ||||| ||||||||| ||||
Db    454 GCCATTGATTCCACTTCTGAAGCCATTTTTGTCACTGCTATGGATACCAATCCGTTGGCA 513

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          || || || || ||  | ||||| |||    | || |  |||      || ||    ||
Db    514 GCTGAGCCGACCGTTGTGATCAATGAACAATCTGAAGCCTTCGTTGCTGGTCTAGATGTG 573

Qy    541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
            |   |    ||||     |||||  |  | ||||||||    |   |||  |   | ||||
Db    574 CTTTCCGCACTGACAACTGGCAAAGTGTACGTTTGTAA---AAAAGGCACCAGCTTGCCA 630

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
```

```
Db    631 CGTTCTCAACAGCCGAATGTTGAAGAACATGTCTTTGACGGCCCACATCCTGCAGGCTTA 690

Qy    661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
            |||||| ||||| |||||| || | || ||  |  ||  |  |    | ||||| ||||
Db    691 GCTGGCACTCACATGCATTTCCTCTATCCAGTAAGTGCAGACCACGTGGCGTGGAGCATC 750

Qy    721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
          || ||||| || || ||  |  | || |   ||||    || |||   ||  ||||| ||
Db    751 AACTATCAGGACGTGATCGCAGTAGGTCAACTGTTCCTGACGGGCGAGCTATACACCCAG 810

Qy    781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          || ||| || | || |  ||  |    || |||||||| ||  |  | || || ||  ||
Db    811 CGTGTGGTTTCTCTTGCCGGCCCGGTGGTGAACAAACCTCGTTTGGTTCGCACTGTCATG 870

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
          || ||   | ||  |||| |    |  ||||| |  |        ||    ||| | ||
Db    871 GGCGCATCACTAGAGCAACTGGTGGACAGCGAAATCATGCCTGGCGAAGTACGCATCATC 930

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          || || || ||  ||   ||  ||| |    ||  | |||| |||| | || || |||
Db    931 TCTGGCTCTGTGCTGTCTGGTACGAAAGCCACTGGACCACACGCTTATCTTGGTCGTTAC 990

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          |||   ||| | || ||| |    ||||| ||    ||||| ||||| || ||||    |
Db    991 CACCTACAGGTCTCTGTTCTTCGTGAAGGACGTGATAAAGAACTGTTTGGTTGGGCGATG 1050

Qy   1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
          |||     | |||| | |||  | || ||  |   ||| || ||  |  | ||   | ||
Db   1051 CCGGGTAAGAACAAGTTCTCTGTGACTCGCTCTTTCCTTGGTCACCTATTCAAGGGCCAA 1110

Qy   1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
            | | |||  | | |||| ||   || ||  | || ||  | ||||| || || ||||
Db   1111 GTGTACAACATGACCACAACCACTAATGGTAGTGATCGATCTATGGTTCCAATCGGTAAC 1170

Qy   1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
          || |||   || |||||  | |||||   ||| ||| || |  | || ||| |    |
Db   1171 TATGAGAAGGTCATGCCACTAGATATGGAGCCAACCTTGTTGCTTCGTGATCTGTGTGCT 1230

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
          || || |  |||||||||     || |||   || ||  |||| |||||||| || ||
Db   1231 GGTGACAGTGACAGCGCGGTTCGCTTAGGTGCTTTAGAGCTGGATGAAGAAGATCTAGCA 1290

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
            |||| |  ||||| ||||| ||||||||||| |||||  | |||  ||||  |    |
Db   1291 CTGTGTACGTTCGTTTGCCCAGGCAAATACGAGTACGGTCAGCTCCTGCGTGAGTGCTTA 1350

Qy   1321 gaaaccattgagaaggaaggctga 1344
          || |  || ||||| ||||| | |
Db   1351 GACAAGATCGAGAAAGAAGGGTAA 1374
```

```
RESULT  12
AE004300/c
LOCUS       AE004300               13382 bp    DNA     linear   BCT 31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 208 of 251 of the complete
            chromosome.
ACCESSION   AE004300 AE003852
VERSION     AE004300.1  GI:9656850
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 13382)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
```

```
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 13382)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .13382
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(154. .1236)
                     /gene="VC2287"
     CDS             complement(154. .1236)
                     /gene="VC2287"
                     /note="similar to GB:U00096 SP:Q47155 PID:1552799
                     PID:1786425; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-damage-inducible protein P"
                     /protein_id="AAF95431.1"
                     /db_xref="GI:9656851"
                     /translation="MQDRIRKIIHVDMDCFFAAVEMRDNPAYREIALAVGGHEKQRGV
                     ISTCNYQARKFGVRSAMPTAQALKLCPQLHVVPGRMSVYKSVSQQIQTIFQRYTSLIE
                     PLSLDEAYLDVSESTAYQGSATLIAQAIRRDIWQELNLTASAGVAPIKFLAKVASDLN
                     KPDGLYVVTPDKVQEMVDSLPLEKIPGVGKVALEKLHQAGLYVGADVRRADYRKLLHQ
                     FGRLGASLWKKSHGIDEREVVTERERKSVGVEYTFSQNISTFQECWQVIEQKLYPELD
                     ARLSRAHPQRGIIKQGIKVKFADFQQTTIEHVHPALELDYFHELLEQVLTRQQGREIR
                     LLGLSVMLKPELQMKQLSMFPSDGWQ"
     gene            complement(1345. .1578)
                     /gene="VC2288"
     CDS             complement(1345. .1578)
                     /gene="VC2288"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF95432.1"
                     /db_xref="GI:9656852"
                     /translation="MNTFLITFGFFLAVIAAMAVGYIFQRKVVKGSCGGLGAVGIEKV
                     CNCPEPCDARKKREAKAAARAERLAAWEKDRIA"
     gene            complement(1595. .2698)
                     /gene="VC2289"
     CDS             complement(1595. .2698)
                     /gene="VC2289"
                     /note="similar to GB:L42023 SP:P44550 PID:1003244
                     PID:1222087 PID:1204428; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="thiamin biosynthesis lipoprotein ApbE"
                     /protein_id="AAF95433.1"
                     /db_xref="GI:9656853"
                     /translation="MFFTNWILKVSMVIWFRVSVFPISIFNYERSKQVRNWLVALASL
                     LLLAGCEKPAEQVHLSGPTMGTTYNIKYIQQPGIADSKTLQTEIDRLLEEVNDQMSTY
                     RKDSELSRFNQHTSSEPFAVSTQTLTVVKEAIRLNGLTEGALDVTVGPLVNLWGFGPE
                     ARPDVVPTDEELNARRAITGIEHLTIEGNTLSKDIPELYVDLSTIAKGWGVDVVADYL
                     QSQGIENYMVEIGGEIRLKGLNRDGVPWRIAIEKPSVDQRSVQEIIEPGDYAIATSGD
                     YRNYFEQDGVRYSHIIDPTTGRPINNRVVSVTVLDKSCMTADGLATGLMVMGEERGMA
                     VAEANQIPVLMIVKTDDGFKEYASSSFKPFLSK"
     gene            complement(2745. .3971)
                     /gene="VC2290"
     CDS             complement(2745. .3971)
                     /gene="VC2290"
                     /note="similar to PID:893416; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, beta subunit"
                     /protein_id="AAF95434.1"
                     /db_xref="GI:9656854"
                     /translation="MSTIIFGVVMFTLIILALVLVILFAKSKLVPTGDITISINGDPE
                     KAIVTQPGGKLLTALAGAGVFVSSACGGGSCGQCRVKIKSGGGDILPTELDHISKGE
                     AREGERLACQVAVKADMDLELPEEIFGVKKWECTVISNDNKATFIKELKLAIPDGESV
                     PFRAGGYIQIEAPAHHVKYADFDVPEKYRGDWDKFNLFRYESKVDEPIIRAYSMANYP
                     EEFGIIMLNVRIATPPPNNPNVPPGQMSSYIWSLKAGDKCTISGPFGEFFAKDTDAEM
                     VFIGGGAGMAPMRSHIFDQLKRLKSKRKMSYWYGARSKREMFYVEDFDGLAAENDNFV
                     WHCALSDPQPEDNWTGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPMMNAAVINMLK
                     NLGVEEENILLDDFGG"
     gene            complement(4005. .4601)
                     /gene="VC2291"
     CDS             complement(4005. .4601)
                     /gene="VC2291"
                     /note="similar to PID:893416 PID:893415; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, hydrophobic membrane protein NqrE"
                     /protein_id="AAF95435.1"
                     /db_xref="GI:9656855"
                     /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGIA
                     VIVVLTISVPVNNLVYNLVLKPDALVEGVDLSFLNFITFIGVIAALVQILEMILDRFF
                     PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYSFAESVVYGFGSGVGWMLAIVALAGI
                     REKMKYSDVPPGLRGLGITFITAGLMALGFMSFSGVQL"
     gene            complement(4609. .5241)
                     /gene="VC2292"
     CDS             complement(4609. .5241)
                     /gene="VC2292"
                     /note="similar to PID:663269 PID:663273 PID:893414;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, hydrophobic membrane protein NqrD"
                     /protein_id="AAF95436.1"
                     /db_xref="GI:9656856"
                     /translation="MSSAKELKKSVLAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVMFVTALSNFFVSLIRNHIPNSVRIIVQMAIIASLVIVVDQILKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSEPIPSFIDGIGNGLGYGFVLMTVGFFRELLGSGKL
                     FGLEVLPLISNGGWYQPNGLMLLAPSAFFLIGFMIWAIRTFKPEQVEAKE"
     gene            complement(5241. .6014)
                     /gene="VC2293"
     CDS             complement(5241. .6014)
                     /gene="VC2293"
                     /note="similar to PID:663269 PID:663272 PID:893413;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, gamma subunit"
                     /protein_id="AAF95437.1"
                     /db_xref="GI:9656857"
                     /translation="MASNNDSIKKTLFVVIALSLVCSIIVSAAAVGLRDKQKENAALD
                     KQSKILQVAGIEAKGSKQIVELFNKSIEPRLVDFNTGDFVEGDAANYDQRKAAKEASE
                     SIKLTAEQDKAKIQRRANVGVVYLVKDGDKTSKVILPVHGNGLWSMMYAFVAVETDGN
                     TVSGLTYYEQGETPGLGGEVENPAWRAQWVGKKLFDENHKPAIKIVKGGAPQGSEHGV
                     DGLSGATLTSNGVQNTFDFWLGDMGFGPFLTKVRDGGLN"
     gene            complement(6001. .7284)
                     /gene="VC2294"
     CDS             complement(6001. .7284)
                     /gene="VC2294"
                     /note="similar to PID:663269 PID:663271; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, hydrophobic membrane protein NqrB"
                     /protein_id="AAF95438.1"
                     /db_xref="GI:9656858"
                     /translation="MLRQDRERRVISMGLKKFLEDIEHHFEPGGKHEKWFALYEAAAT
                     LFYTPGLVTKRSSHVRDSVDLKRIMIMVWLAVFPAMFWGMYNAGGQAIAALNHLYSGD
                     QLAAIVAGNWHYWLTEMLGGTMSSDAGWGSKMLLGATYFLPIYATVFIVGGFWEVLFC
                     MVRKHEVNEGFFVTSILFALIVPPTLPLWQAALGITFGVVVAKEVFGGTGRNFLNPAL
                     AGRAFLFFAYPAQISGDLVWTAADGYSGATALSQWAQGGAGALINNATGQTITWMDAF
                     IGNIPGSIGEVSTLALMIGAAFIVYMGIASWRIIGGVMIGMILLSTLFNVIGSDTNAM
                     FNMPWHWHLVLGGFAFGMFFMATDPVSASFTNSGKWAYGILIGVMCVLIRVVNPAYPE
                     GMMLAILFANLFAPLFDHVVVERNIKRRLARYGKQ"
     gene            complement(7254. .8642)
                     /gene="VC2295"
     CDS             complement(7254. .8642)
                     /gene="VC2295"
                     /note="similar to PID:663269 PID:663270; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, alpha subunit"
                     /protein_id="AAF95439.1"
                     /db_xref="GI:9656859"
                     /translation="MRNERIFFPIKVVQVRMITIKKGLDLPIAGTPSQVISDGKAIKK
                     VALLGEEYVGMRPTMHVRVGDEVKKAQILFEDKKNPGVKFTSPVSGKVVEINRGAKRV
                     LQSVVIEVAGDDQVTFDKFEANQLASLNRDAIKTQLVESGLWTAFRTRPFSKVPAIDS
                     TSEAIFVTAMDTNPLAAEPTVVINEQSEAFVAGLDVLSALTTGKVYVCKKGTSLPRSQ
                     QPNVEEHVFDGPHPAGLAGTHMHFLYPVSADHVAWSINYQDVIAVGQLFLTGELYTQR
                     VVSLAGPVVNKPRLVRTVMGASLEQLVDSEIMPGEVRIISGSVLSGTKATGPHAYLGR
                     YHLQVSVLREGRDKELFGWAMPGKNKFSVTRSFLGHLFKGQVYNMTTTTNGSDRSMVP
                     IGNYEKVMPLDMEPTLLLRDLCAGDSDSAVRLGALELDEEDLALCTFVCPGKYEYGQL
                     LRECLDKIEKEG"
     gene            complement(9079. .9399)
                     /gene="VC2296"
```

Query Match 36.4%; Score 489.6; DB 1; Length 13382;
Best Local Similarity 60.8%; Pred. No. 2.5e-109;
Matches 817; Conservative 0; Mismatches 524; Indels 3; Gaps 1;

```
Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
Db    8594 ATGATTACAATAAAAAAGGGATTGGACCTTCCTATCGCAGGAACTCCTTCCCAGGTGATT 8535

Qy      61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
Db    8534 AGTGATGGTAAAGCCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 8475

Qy     121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
Db    8474 CCTACTATGCATGTCCGCGTAGGTGATGAAGTAAAAAAAGCTCAGATTCTTTTTGAAGAT 8415

Qy     181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
Db    8414 AAGAAGAATCCGGGTGTGAAATTTACTTCACCAGTCAGCGGTAAAGTGGTTGAAATCAAT 8355

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
Db    8354 CGTGGTGCAAAGCGCGTTCTTCAGTCTGTAGTGATTGAAGTGGCGGGTGATGACCAAGTG 8295

Qy     301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
Db    8294 ACCTTCGACAAGTTTGAAGCCAATCAATTAGCCAGTTTGAATCGCGATGCGATCAAAACT 8235

Qy     361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
Db    8234 CAGCTGGTTGAGTCTGGTCTTTGGACTGCTTTCCGTACTCGTCCGTTCAGCAAGGTTCCA 8175

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
Db    8174 GCCATTGATTCCACTTCTGAAGCCATTTTTGTCACTGCTATGGATACCAATCCGTTGGCA 8115

Qy     481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
Db    8114 GCTGAGCCGACCGTTGTGATCAATGAACAATCTGAAGCCTTCGTTGCTGGTCTAGATGTG 8055

Qy     541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
Db    8054 CTTTCCGCACTGACAACTGGCAAAGTGTACGTTTGTAA---AAAAGGCACCAGCTTGCCA 7998

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
Db    7997 CGTTCTCAACAGCCGAATGTTGAAGAACATGTCTTTGACGGCCCACATCCTGCAGGCTTA 7938

Qy     661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
Db    7937 GCTGGCACTCACATGCATTTCCTCTATCCAGTAAGTGCAGACCACGTGGCGTGGAGCATC 7878

Qy     721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
Db    7877 AACTATCAGGACGTGATCGCAGTAGGTCAACTGTTCCTGACGGGCGAGCTATACACCCAG 7818

Qy     781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
Db    7817 CGTGTGGTTTCTCTTGCCGGCCCGGTGGTGAACAAACCTCGTTTGGTTCGCACTGTCATG 7758

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
Db    7757 GGCGCATCACTAGAGCAACTGGTGGACAGCGAAATCATGCCTGGCGAAGTACGCATCATC 7698

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
Db    7697 TCTGGCTCTGTGCTGTCTGGTACGAAAGCCACTGGACCACACGCTTATCTTGGTCGTTAC 7638

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db    7637 CACCTACAGGTCTCTGTTCTTCGTGAAGGACGTGATAAAGAACTGTTTGGTTGGGCGATG 7578

Qy    1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
Db    7577 CCGGGTAAGAACAAGTTCTCTGTGACTCGCTCTTTCCTTGGTCACCTATTCAAGGGCCAA 7518

Qy    1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
Db    7517 GTGTACAACATGACCACAACCACTAATGGTAGTGATCGATCTATGGTTCCAATCGGTAAC 7458

Qy    1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
Db    7457 TATGAGAAGGTCATGCCACTAGATATGGAGCCAACCTTGTTGCTTCGTGATCTGTGTGCT 7398

Qy    1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
Db    7397 GGTGACAGTGACAGCGCGGTTCGCTTAGGTGCTTTAGAGCTGGATGAAGAAGATCTAGCA 7338

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
Db    7337 CTGTGTACGTTCGTTTGCCCAGGCAAATACGAGTACGGTCAGCTCCTGCGTGAGTGCTTA 7278

Qy    1321 gaaaccattgagaaggaaggctga 1344
Db    7277 GACAAGATCGAGAAAGAAGGGTAA 7254
```

```
RESULT 13
AE004724/c
LOCUS       AE004724               16903 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 285 of 529 of the complete
            genome.
ACCESSION   AE004724 AE004091
VERSION     AE004724.1  GI:9949083
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 16903)
```

```
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 16903)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .16903
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            complement(67. .2292)
                     /gene="PA2984"
     CDS             complement(67. .2292)
                     /gene="PA2984"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06372.1"
                     /db_xref="GI:9949084"
                     /translation="MRLGLPALAAGLLALRFLAELPSVGWLLGMAAAGLLLLFGRLYW
                     LGLFLLGFAWACGSAQVALDGRLAAELDGRTLWLEGRVVDLPASGEGVTRFRLEEVSA
                     ARGVLLPERLRLSWRGAPPLMGGERWRLAVTLKRPRGLVNAAGFDYEAWLLAQRVGAV
                     GTVKAGQRLQPASGLDAWRDAWRQRLLAVDAQGRGAALAALVLGDGSGLSAAEWRMFQ
                     DTGTVHLMVISGSHISLLAGMLYGLVAALHRLGCWPRRLPWLPCACALGLLGAWSYGL
                     MAGFEVPVRRACLMVSLALVWRLRFRHLGLLTPLLATLVLVLLVEPLASLQAGFWLSF
                     LAVALLLWLFAGRLGRWRWWAAWGRAQWGMALGLAMPLLALGLPLSLSGAPANLLAVP
                     WIELLVVPPALAGSLLLWVPGLGEGLLRAAGVSLDLLFRALGLLAGLAEAWQPPAAAA
                     WSVALGMLGALCWLAPAGLPLRALGAALLLPALLPSSPPVEWGRAEVRVLDVGQGLAV
                     LVRTREHVLLYDSGARQGAFDMGERVVVPVLRSLDLRRLDGLLLSHADNDHAGGAPTV
                     ASRFPPVWLVSGEPARLPSPLFADSCDERSWSWDGVVFEQWAWAQAGDSNDRSCVLRV
                     EADGEVLLLTGDISRAAEHAWLARQADPRVDWLLAPHHGSRSSSGVAFVLRTRPRHVL
                     VSRGWRNAFGHPHGEVMERYARVAAQVHDTARDGALTFLLGSRGGARRERDSAHFWRE
                     K"
     gene            2423. .2956
                     /gene="PA2985"
     CDS             2423. .2956
                     /gene="PA2985"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06373.1"
                     /db_xref="GI:9949085"
                     /translation="MPRRLFKRYMPDPESIRSQKSLRFLGKLTHDPNLWHLNRHSVSR
                     AMAAGLFAAFIPMPMQMLLAAGLAVWIRSNLPISVGLVWLTNPITMPPVFYCTYKLGA
                     WLMQVPPRTLPDHLTWEWISQELSSLWQPFLLGSVVCGIIAAILGYALTMLYWRWWVQ
                     RSWSRRCRQRRERVAAK"
     gene            complement(3044. .4345)
                     /gene="PA2986"
     CDS             complement(3044. .4345)
                     /gene="PA2986"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG06374.1"
                     /db_xref="GI:9949086"
                     /translation="MPAVRSRPANPPSTRCIRDMFRPLPFFIGLRYTRAKRRNHYISF
                     ISLTSMIGLALGVLAMIIVLSVMNGFQKEMRTRILGMVPHATISAAQPLDDWQTVANA
                     ALRHREVVGAAPFAELQGMLSYKGNMLPVLVNGIDPQEERKVSIVGEHIVQGSLDDLK
                     PGEFGIVLGEITARRFHVNVGDKLTLIVPEATSAPGGITPRMQRFTIVALFKVGAELD
                     NSLALIDIADAGQLLRLQPGQVPSVRLELKDLYQSPQVAAKVVKELGQGFRSSDWTRT
                     QGSLFNAMKMEKTMIGLLLLLIIAVAAFNIIATLIMVVADKRTDIAILRTLGATPRQI
                     MAIFMVQGTVIGVIGTVIGGVLGVFAALNITGMIDRIERLVGHKVFSSDVYFINYLPS
                     DLQVLDVVLICSAALLMSFLATLYPSWRAARTQPAESLRYE"
     gene            complement(4358. .5041)
                     /gene="PA2987"
     CDS             complement(4358. .5041)
                     /gene="PA2987"
                     /codon_start=1
                     /transl_table=11
                     /product="probable ATP-binding component of ABC
                     transporter"
                     /protein_id="AAG06375.1"
                     /db_xref="GI:9949087"
                     /translation="MNDKSVLSCRNLSKSYDEGPQSVQVLSGVELNLLPGERVAIVGS
                     SGSGKSTLLNMLGGLDTPSAGSVWLAGEELSALNETARGLLRNRALGFVYQFHHLLPE
                     FTALENVCMPLLIGRTPIAEARQRAAELLERVGLGHRLSHKPAELSGGERQRVAIARA
                     LVNRPQLVLLDEPTGNLDQHTAQGIQELMLELSRSLQTSFLVVTHDLQLAGHMDRILR
                     LEEGRLIAA"
     gene            complement(5034. .6284)
                     /gene="PA2988"
     CDS             complement(5034. .6284)
                     /gene="PA2988"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG06376.1"
                     /db_xref="GI:9949088"
                     /translation="MFRPLSVFIGTRYTRAKRRSHFVSFISLTSMIGLALGVLVMIVV
                     LSVMNGFDREMRTRILGMVPQATVESYQPIDDWRALAEKVKAHEHVTAVAPFTQMQGM
                     LSANGAVQTVMVDAVDPLEEAKVSIIPDFFKQGSLAELKAGGFGIVIGQLAAQKLGVG
                     IGDKVTLILPEVAVTPAGVFPRMKRFTVVGTFRVGAGELDGGLSLIHLEDAARLQRWK
                     TNQVQGLRLKLDDLFQAPRVAWEIARTLTDNDFYARDWTRSHGNLYQAIRMEKTMIGL
                     LLLLIVAVAAFNIISTLVMVVTDKKSDIAILRTLGATPGQIMATFMVQGTVIGVIGTL
                     VGGVLGVVAALNVSAWISALEKLLGHQFLASDVYFIDYLPSQLMLDDVVLVCGAALVL
                     SFFATLYPAWRAARTQPAEALRYE"
     gene            6222. .6986
                     /gene="PA2989"
     CDS             6222. .6986
                     /gene="PA2989"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06377.1"
                     /db_xref="GI:9949089"
                     /translation="MAATFGPRIARADENGQGSEHVALSWQKMDASGWRSGEASPAAL
                     HSDHHCRRGFAMSTRDADDRREYYRIEDTLALEFRPAGAVEDGLESEPGATLFSLLSD
                     LHLMDYESQHLLRHISERDRTLANYLKVMNKRIDLLGQVMVQSLLKEIGEPRKVSLSE
                     GGVSFRHDRALPVGQLLVLRMVLLPQGFGLELRARVIHAQPHDDEFEIGTEFEALSDA
                     QRQLLARHILQKQAQQRRLARAGQGPLGEPGQPSST"
     gene            7318. .8040
                     /gene="PA2990"
     CDS             7318. .8040
                     /gene="PA2990"
                     /codon_start=1
                     /transl_table=11
                     /product="probable phosphodiesterase"
                     /protein_id="AAG06378.1"
                     /db_xref="GI:9949090"
                     /translation="MTLIYGHRGAKGEAPENTLASFQRCLEHGVNRCELDLHLSRDGE
                     LMVIHDPTLKRTTGRRGKVAEHDAATLATYDARKGGPGWVRPTPIPRLEELFEKCPFE
                     HWQLEVKSASRERAARTVTRIRELTERFGNRDKVTVTSGSRTVLKALNELTPELSRGL
                     VAEYAWLDPLKVAQHYGCALLALNWTLCTPERLLKAQRQGLHVSVWTVNEPALMRRLA
                     DFGVDSLITDFPGLASATLGNR"
     gene            complement(8083. .9477)
                     /gene="sth"
                     /note="PA2991"
     CDS             complement(8083. .9477)
                     /gene="sth"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="soluble pyridine nucleotide transhydrogenase"
                     /protein_id="AAG06379.1"
                     /db_xref="GI:9949091"
                     /translation="MAVYNYDVVILGTGPAGEGAAMNASKYGRKLAVVDSRRVVGGNC
                     THLGTIPSKALRHSVKQIIEFNTNPMFRQIGEPRWFSFPDVLKSADRVISKQVASRTG
                     YYARNRIDMFTGTASFVDERTVEVVTPSGAVERLVADQFVIATGSRPYRPSDINFNHP
                     RVYDSDTILSLSHTPRRLIIYGAGVIGCEYASIFSGLGVLVDLIDTRDQLLSFLDDEI
                     SDALSYHLRNNNVLIRHNEEYERVEGLDNGVILHLKSGKKIKADALLWCNGRTGNTDK
                     LGLENVGIKVNSRGQIEVDENYRTSVSNIFAAGDVIGWPSLASAAYDQGRSAAGNIVE
                     SDSWRFVNDVPTGIYTIPEISSIGKNESELTAAKIPYEVGKAFFKGMARAQISNEPVG
                     MLKILFHRETLEILGVHCFGDQASEIVHIGQAIMNQPGELNTLKYFVNTTFNYPTMAE
                     AYRVAAFDGLNRLF"
     gene            complement(9656. .9883)
                     /gene="PA2992"
     CDS             complement(9656. .9883)
                     /gene="PA2992"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06380.1"
                     /db_xref="GI:9949092"
                     /translation="MTWLLVFVIMLLVVFGMSIGVIMGRKPIAGSCGGIANLGIEKEC
                     SICGGDRQKCEEANRDESVVPGKLAYDATKR"
     gene            complement(9880. .10908)
                     /gene="PA2993"
     CDS             complement(9880. .10908)
                     /gene="PA2993"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG06381.1"
                     /db_xref="GI:9949093"
                     /translation="MAERFRPLAMAGLAALALVLAGCGDTLESFGGPTMGSTYSIKYV
                     RGASAPDVQTAKAAVEAILAEVDRQMSTYRDDSLVSRFNALPAQSCMELPLPMLELLR
                     YGGELSEQSQGAFDMTVEPLMNLWGFGPQARVEKVPSAEQIAAVRRDVGHRHLRIDGQ
                     RLCKDAAVQLDFDSIAAGYTVDAVGERLKELGVRSYLAEITGELKAEGRKPDGTPWRI
                     AIEAPREGQRVAQQVLALDGYGVSTSGDYRNYFEENGQRYSHTFDPRTGAPIDHHLAS
                     VTVIDPSTRNADGLSTLLMVLGPEEGYRFAEKHRLAALFVSRQGNGFDSRTTPRYEQL
                     FGNQGERP"
     gene            complement(10901. .12124)
                     /gene="nqrF"
                     /note="PA2994"
     CDS             complement(10901. .12124)

  Query Match             35.5%;  Score 477.2;  DB 1;  Length 16903;
  Best Local Similarity   60.3%;  Pred. No. 2.8e-106;
  Matches  810;  Conservative    0;  Mismatches  528;  Indels    6;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgttt 60
          |||||| || || |||   || ||  ||||||||||||   |||   || || ||   | |
Db  16744 ATGATCAAGATAAAACGTGGCCTGGACCTGCCCATCAGCGGCGCCCCCGAACAGCGCATC 16685

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           | |  |  |   |  |   |   || |||||| | ||    || ||   ||| |||
Db  16684 GAGGCAGCACGCCCGGTACGCAGCGTGGCGTTGATCGGTTTCGACTACCACGGCATGAAG 16625

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||  | |||   ||  ||| |||||||    || ||    || || ||||||||    |||
Db  16624 CCGACCATGGCGGTGCAGGTAGGCGACCGGGTGAAGCTGGGGCAGGTGCTGTTCACCGAC 16565

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          || ||||||||   |||    | ||| || || |   | ||     ||   || || ||
Db  16564 AAGAAGAATCCCTCCGTCAGCTATACCGCACCCGGCGCCGGTGTGGTCAGTGCCATCCAT 16505

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          |||||||| |||||||| || || || || || || |   | || ||| |||| || | |
Db  16504 CGTGGCGAGAAGCGCGTCCTGCAATCGGTGGTCATCGATCTCGACGGCGACGAGCAACTG 16445

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
          |||||  |  || ||  |  ||||    || || |   | |||| |||  | || |||  |
Db  16444 GAGTTCGCCCGTTATCCCGCTGACAAACTCGCCACGCTGAGCGCCGAGCAGGTCCGCGAC 16385

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
          |||||| | || || ||  ||||||| |||||||| ||||| ||||||||||  | ||
Db  16384 AACCTGTTGCAGTCGGGCCTGTGGACCGCGCTGCGGACCCGGCCGTTCAGCAAGGTGCCC 16325

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          |    |||  ||  ||     | || |||||||| ||| || || ||| | ||||||||
Db  16324 GATCCCGAGTCCAGCCCCAGTTCGATTTTCGTCACTGCCATCGATACCCAGCCGCTGGCC 16265

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          ||||| ||   |||  |||||   ||    | |||||  ||| |   ||| |||  |||
Db  16264 GCCGATCCGCAGGTGGTTATCGCGGAGCAGGGCGAGGCCTTCCAGGCCGGGCTGACGGTG 16205

Qy    541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
            | | ||  || |     |    |     | || |||||  |  ||||      | ||
Db  16204 CTCGGGCGCCTGGC------CCGGGTGTTCCTGTGCAAGGCGGAAGGCGTTTCGCTACCC 16151

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
            ||         ||  | |     | || | ||||  ||| |||||||| || ||  ||
Db  16150 GGCGAGGCGCTCTCCGGCGTGACCGCGCAAGCATTCTCCGGTCCGCATCCGGCGGGCCTG 16091

Qy    661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
            ||||| || || || ||| | || ||||| || ||    || | ||||||||  | |
Db  16090 CCCGGCACCCATATCCACTTCCTTGACCCGGTGGGTGCCGGCAAGAGCGTGTGGAACCTG 16031

Qy    721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
          || || || || || ||  |||| |||    ||||  | || || |   ||  ||||||
Db  16030 AACTACCAGGACGTGATCGCCATCGGCAAGCTGTTCACGACCGGGCAATTGTGGACCGAG 15971

Qy    781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          || ||||| ||||| |  ||| |    ||  | || ||||| ||  |||| ||| |||
Db  15970 CGGGTGATCGCCCTGGCCGGTCCGGTGGTGGAAAAGCCGCGTCTGGTGCGCACCCGCTTG 15911

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
          || || ||  |    || |  | || ||| | |||          |||||||  ||||
Db  15910 GGCGCCAACCTCGACGAACTGGCCGCCGGCCAGTTGCAGCCGGGCAACAACCGGCTGATC 15851

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          ||||| || ||  ||  ||| |   | || ||||| ||| ||  |||| ||||||
Db  15850 TCCGGCTCCGTGCTGGGCGGTCGCACCGCCCATGGCGCCTACGCCTACCTGGGGCGCTAC 15791

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          |||   ||| | |||  | || | || |||  |      ||| |  ||  ||  |   |
Db  15790 CACCTGCAGTTGTCCTGTCTCAAGGAGGGCGACCAGCGCGAGTTCCTCCACTACCTGCGC 15731

Qy   1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
            |   | ||| ||  | ||  |             | ||  |    | ||      |||
Db  15730 GCCGGGGTGGAAAAGCATTCGCTGCTCAACGTGTTCGTCTCCAGGCTGCTCGGCGGCAAG 15671

Qy   1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
          | ||||   |||| |||   |  ||||||| ||   || || |||||||| | ||||
Db  15670 CGCTTCGCCTTCACCACCAGCACCAACGGCAGCCCACGGGCGATGGTGCCGGTAGGTAAC 15611

Qy   1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
          ||||||   ||||||||  |||||||||||||| || |   |  ||||| |  | ||||||
Db  15610 TACGAGGCGGTGATGCCGCTGGATATCCTGCCGACTCAATTGCTGCGCTACCTGATCGTC 15551

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
          ||||| || || |  || |||    | ||   | ||||| | |||||||| ||||| ||
Db  15550 GGCGACACGGAAATGGCCCAGAAGCTCGGCGCCCTGGAACTCGACGAAGAGGACCTGGCG 15491

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          |||||||| |||||| | ||||| ||||||||||||||| |  |||| |    |||
Db  15490 CTGTGCAGCTACGTCTGCGCCGGCAAGTACGAATACGGCCCGATTCTGCGGGACAATCTG 15431

Qy   1321 gaaaccattgagaaggaaggctga 1344
          | |  ||| ||| |||| ||||||
Db  15430 GCACGCATCGAGCAGGAGGGCTGA 15407


RESULT 14
VANQRBOL
LOCUS       VANQRBOL                5078 bp    DNA     linear   BCT 15-JUN-1999
DEFINITION  V.alginolyticus bolA, nqrA, nqrB, nqrC, nqrD and nqrE (partial)
            genes.
```

```
ACCESSION   Z37111
VERSION     Z37111.1  GI:663268
KEYWORDS    bolA gene; BolA protein; NADH ubiquinone oxidoreductase;
            NADH:ubiquinone oxidoreductase subunit A; NADH:ubiquinone
            oxidoreductase subunit B; nqrA gene; nqrB gene; nqrC gene; nqrD
            gene; nqrE gene.
SOURCE      Vibrio alginolyticus.
  ORGANISM  Vibrio alginolyticus
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 2384)
  AUTHORS   Beattie,P., Tan,K., Bourne,R.M., Leach,D., Rich,P.R. and Ward,F.B.
  TITLE     Cloning and sequencing of four structural genes for the
            Na(+)-translocating NADH-ubiquinone oxidoreductase of Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 356 (2-3), 333-338 (1994)
  MEDLINE   95104445
REFERENCE   2  (bases 1 to 5078)
  AUTHORS   Ward,B.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1994) Ward B. F., Edinburgh University, Institute
            for Cell and Molecular Biology, Mayfield Rd, Edinburgh, Scotland,
            EH9 3JR
FEATURES             Location/Qualifiers
     source          1. .5078
                     /organism="Vibrio alginolyticus"
                     /db_xref="taxon:663"
                     /clone="10-2"
                     /clone_lib="Lambda library"
     gene            26. .334
                     /gene="bolA"
     CDS             26. .334
                     /gene="bolA"
                     /function="Cell division"
                     /codon_start=1
                     /transl_table=11
                     /product="BolA"
                     /protein_id="CAA85475.1"
                     /db_xref="GI:663269"
                     /db_xref="SWISS-PROT:Q56585"
                     /translation="MIQEIIEKKLHSELQPSYLKVINESYMHNVPPGSESHFKVIVVS
                     DSFAGQRLIGRHRQVHQILADELDNHIHALAIHTYTDEEWKREQDGAPDSPMCMGGGH
                     "
     gene            808. .2148
                     /gene="nqrA"
     CDS             808. .2148
                     /gene="nqrA"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="NADH:ubiquinone oxidoreductase subunit A"
                     /protein_id="CAA85476.1"
                     /db_xref="GI:663270"
                     /db_xref="SPTREMBL:Q56586"
                     /translation="MITIKKGLDLPIAGTPSQVINDGKTIKKVALLGEEYVGMRPTMH
                     VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEVNRGAKRVLQSVVIEVAGEEQVTF
                     DKFEAAQLSGLDREVIKTQLVDSGLWTALRTRPFSKVPAIESSTKAIFVTAMDTNPLA
                     AKPELIINEQQEAFIAGLDILSALTEGKVYVCKSGTSLPRSSQSNVEEHVFDGPHPAG
                     LAGTHMHFLYPVNAENVAWSINYQDVIAFGKLFLTGELYTDRVVSLAGPVVNNPRLVR
                     TVIGASLDDLTDNELMPGEVRVISGSVLTGTHATGPHAYLGRYHQQVSVLREGREKEL
                     FGWAMPGKNKFSVTRSFLGHVFKGQLFNMTTTTNGSDRSMVPIGNYERVMPLDMEPTL
                     LLRDLCAGDTDSAQALGALELDEEDLALCTFVCPGKYEYGTLLRECLDTIEKEG"
     gene            2118. .3398
                     /gene="nqrB"
     CDS             2118. .3398
                     /gene="nqrB"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase subunit B"
                     /protein_id="CAA85477.1"
                     /db_xref="GI:663271"
                     /db_xref="SPTREMBL:Q56587"
                     /translation="MPRYYREGRVIFMALKKFLEDIEHHFEPGGKHEKWFALYEAVAT
                     VFYTPGIVTNKSSHVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGD
                     QLATVISGNWHYWLTEMLGGTIAADAGVGSKMLLGATYFLPIYATVFLVGGFWEVLFC
                     MVRKHEVNEGFFVTSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPAL
                     AGRAFLFFAYPAQISGDVVWTAVDGFSGATALSQWAQGGNGALVNTVTGSPITWMDAF
                     IGNIPGSIGEVSTLALMIGAAMIVYMRIASGRIIAGVMIGMIAVSTLFNVIGSDTNPM
                     FNMPWWHWLVLGGFAFGMFFMATDPVSASFTNKGKWWYGILIGAMCVMIRVVNPAYPE
                     GMMLAILFANLFAPLFDHVVIEKNIKRRLARYGK"
     gene            3388. .4158
                     /gene="nqrC"
     CDS             3388. .4158
                     /gene="nqrC"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="NADH:uniquinone oxidoreductase"
                     /protein_id="CAA85478.1"
                     /db_xref="GI:663272"
                     /db_xref="SPTREMBL:Q56582"
                     /translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDKQKANAVLD
                     KQSKIVEVAGIDANGKKVPELFAEYIEPRLVDLETGNFTEGNASTYDQREASKDAERS
                     IALTPEEDVADIRRRANTAVVYLVKDQDEVQKVILPMHGKGLWSMMYAFVAVETDGNT
                     VSAITYYEQGETPGLGGEVENPSRRDQFIGKKLYNEDHQPAIKVVKGGAPQGSEHGVD
                     GLSGATLTSNGVQHTFDFWLGDKGFGPFLAKVRDGELN"
     gene            4158. .4790
                     /gene="nqrD"
     CDS             4158. .4790
                     /gene="nqrD"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:uniquinone oxidoreductase"
                     /protein_id="CAA85479.1"
                     /db_xref="GI:663273"
                     /db_xref="SPTREMBL:Q57095"
                     /translation="MSSAQNVKKSILAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                     FGLEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
     gene            4798. .5078
                     /gene="NqrE"
     CDS             4798. .>5078
                     /gene="NqrE"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase"
                     /protein_id="CAA85480.1"
                     /db_xref="GI:663274"
                     /db_xref="SWISS-PROT:Q56589"
                     /translation="MEHYISLLVKSISSKHALSFFLGMCTFLAVSKKVKTSFGLGVAV
                     VVVLTIAVPVNNLVYNLVLRENALVEGVDLSFLNFITFIGVIAALVQIL"
BASE COUNT     1366 a   1068 c   1207 g   1437 t
ORIGIN


  Query Match           35.4%;  Score 475.2;  DB 1;  Length 5078;
  Best Local Similarity   60.1%;  Pred. No. 8.1e-106;
  Matches  808;  Conservative    0;  Mismatches  533;  Indels    3;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgttt 60
          ||||||| ||| || || ||  |  | || || ||||| || |  ||   || |  ||
Db    808 ATGATTACAATAAAGAAGGGCTTGGATCTTCCTATCGCAGGAACTCCATCCCAGGTGATT 867

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           | || ||   | |||| |   ||||||| |||||||||||||| || |  || |||||
Db    868 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 927

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||  | ||| | |||   |  || ||||  || |||||||  ||||| || ||||||||
Db    928 CCTACCATGCATGTCCGCGTTGGTGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAT 987

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          |||||||| || ||||||   |||||||| || ||  | || ||| |   ||   || ||
Db    988 AAAAAGAACCCTGGCGTGAAATTTACTGCACCAGCAGCCGGTAAAGTGATCGAAGTTAAC 1047

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
```

```
Db   1048 CGTGGCGCTAAACGTGTCCTTCAATCTGTAGTGATTGAAGTGGCAGGTGAAGAGCAGGTG 1107

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360

Db   1108 ACATTTGATAAGTTCGAAGCCGCTCAACTTTCAGGTCTAGATCGTGAAGTGATCAAGACT 1167

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420

Db   1168 CAGTTGGTTGACTCTGGCCTATGGACCGCTTTACGTACTCGTCCGTTTAGCAAGGTTCCT 1227

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480

Db   1228 GCAATCGAGTCTTCTACTAAGGCTATTTTCGTAACTGCAATGGATACTAATCCATTAGCA 1287

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540

Db   1288 GCTAAGCCTGAGTTGATCATTAATGAGCAACAAGAAGCATTCATTGCTGGTCTAGACATT 1347

Qy    541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600

Db   1348 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAA---ATCTGGCACTAGTTTGCCA 1404

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660

Db   1405 CGCTCTTCACAGTCTAACGTAGAAGAACACGTCTTCGATGGCCCTCACCCAGCAGGTCTT 1464

Qy    661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720

Db   1465 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCATGGAGCATC 1524

Qy    721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780

Db   1525 AACTACCAAGACGTTATCGCGTTCGGTAAGCTGTTCCTAACTGGTGAGCTATACACAGAT 1584

Qy    781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840

Db   1585 CGTGTTGTGTCTCTAGCTGGTCCAGTTGTGAATAACCCTCGTCTAGTTCGTACCGTAATC 1644

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900

Db   1645 GGTGCAAGCTTGGATGACCTAACAGATAACGAGTTGATGCCTGGCGAAGTTCGTGTGATT 1704

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960

Db   1705 TCTGGTTCTGTACTGACAGGTACACATGCAACTGGTCCACACGCATACCTTGGCCGTTAC 1764

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020

Db   1765 CATCAACAGGTTTCTGTATTACGTGAAGGCCGTGAGAAAGAACTATTTGGCTGGGCAATG 1824

Qy   1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080

Db   1825 CCTGGTAAGAATAAGTTCTCGGTAACTCGCTCTTTCCTTGGTCACGTATTCAAAGGTCAG 1884

Qy   1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140

Db   1885 TTGTTCAATATGACAACGACGACCAATGGTAGTGATCGTTCAATGGTTCCAATTGGCAAC 1944

Qy   1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200

Db   1945 TACGAACGCGTAATGCCTCTAGATATGGAGCCTACTCTGCTGCTTCGTGATCTATGTGCA 2004

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260

Db   2005 GGTGATACAGACAGTGCTCAAGCACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 2064

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320

Db   2065 TTGTGTACCTTTGTATGTCCAGGTAAGTACGAGTACGGTACGTTACTTCGTGAATGCCTA 2124

Qy   1321 gaaaccattgagaaggaaggctga 1344

Db   2125 GATACTATCGAGAAGGAAGGGTAA 2148
```

```
RESULT  15
A96090
LOCUS       A96090                   474 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 123 from Patent WO9924578.
ACCESSION   A96090
VERSION     A96090.1  GI:6779937
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 123 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      122 a    130 c    130 g     91 t      1 others
ORIGIN
```

Query Match 35.2%; Score 473.6; DB 6; Length 474;
Best Local Similarity 99.8%; Pred. No. 1.8e-105;
Matches 473; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60

Db      1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 60

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120

Db     61 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180

Db    121 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240

Db    181 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 240

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300

Db    241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAARGCAACGACGAAATC 300

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360

Db    301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 360

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420

Db    361 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 420

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccg 474

Db    421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG 474
```

```
RESULT  16
AB008030
LOCUS       AB008030                 6908 bp    DNA     circular BCT 15-JUN-1999
DEFINITION  Vibrio alginolyticus genes for Na-translocating NADH-quinone
            reductase complex, nqr operon, complete genome.
ACCESSION   AB008030 D43958 D49364
VERSION     AB008030.1  GI:2558472
KEYWORDS    Nqr1 subunit of Na-translocating NADH-quinone reductase complex
            alpha-subunit; Nqr2 subunit of Na-translocating NADH-quinone
```

```
            reductase complex; Nqr3 subunit of Na-translocating NADH-quinone
            reductase complex gamma-subunit; Nqr4 subunit of Na-translocating
            NADH-quinone reductase complex; Nqr5 subunit of Na-translocating
            NADH-quinone reductase complex; Nqr6 subunit of Na-translocating
            NADH-quinone reductase complex beta-subunit.
SOURCE      Vibrio alginolyticus DNA.
  ORGANISM  Vibrio alginolyticus
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (sites)
  AUTHORS   Hayashi,M., Hirai,K. and Unemoto,T.
  TITLE     Cloning of the Na(+)-translocating NADH-quinone reductase gene from
            the marine bacterium Vibrio alginolyticus and the expression of the
            beta-subunit in Escherichia coli
  JOURNAL   FEBS Lett. 356 (2-3), 330-332 (1994)
  MEDLINE   95104444
REFERENCE   2  (sites)
  AUTHORS   Beattie,P., Tan,K., Bourne,R.M., Leach,D., Rich,P.R. and Ward,F.B.
  TITLE     Cloning and sequencing of four structural genes for the
            Na(+)-translocating NADH-ubiquinone oxidoreductase of Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 356 (2-3), 333-338 (1994)
  MEDLINE   95104445
REFERENCE   3  (sites)
  AUTHORS   Hayashi,M., Hirai,K. and Unemoto,T.
  TITLE     Sequencing and the alignment of structural genes in the nqr operon
            encoding the Na(+)-translocating NADH-quinone reductase from Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 363 (1-2), 75-77 (1995)
  MEDLINE   95246889
REFERENCE   4  (sites)
  AUTHORS   Hayashi,M., Nakayama,Y. and Unemoto,T.
  TITLE     Existence of Na+-translocating NADH-quinone reductase in
            Haemophilus influenzae
  JOURNAL   FEBS Lett. 381 (3), 174-176 (1996)
  MEDLINE   96176316
REFERENCE   5  (sites)
  AUTHORS   Nakayama,Y., Hayashi,M. and Unemoto,T.
  TITLE     Identification of six subunits constituting Na+-translocating
            NADH-quinone reductase from the marine Vibrio alginolyticus
  JOURNAL   FEBS Lett. 422 (2), 240-242 (1998)
  MEDLINE   98149659
REFERENCE   6  (bases 1 to 6908)
  AUTHORS   Hayashi,M., Unemoto,T. and Sugiyama,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-1997) Maki Hayashi, Faculty of Pharmaceutical
            Sciences, Chiba University, Laboratory of Membrane Biochemistry;
            1-33 Yayoi-cho, Inage-ku, Chiba, Chiba 263, Japan
            (E-mail:makiha@p.chiba-u.ac.jp, Tel:043-290-2932, Fax:043-290-3021)
COMMENT     On Oct 22, 1997 this sequence version replaced gi:677955 gi:893412.
            D43958: submitted (17-Dec-1994)
            D49364: submitted (17-Feb-1995).
FEATURES             Location/Qualifiers
     source          1. .6908
                     /organism="Vibrio alginolyticus"
                     /db_xref="taxon:663"
                     /note="There are conflicts between this sequence and
                     GenBank Accession Number:Z37111."
     gene            386. .1726
                     /gene="nqr1"
     CDS             386. .1726
                     /gene="nqr1"
                     /note="contains FMN as a cofactor; nqrA, Nqr1 subunit of
                     Na-NQR complex alpha-subunit"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr1 subunit of Na-translocating NADH-quinone
                     reductase complex alpha-subunit"
                     /protein_id="BAA22910.1"
                     /db_xref="GI:2558473"
                     /translation="MITIKKGLDLPIAGTPSQVINDGKTIKKVALLGEEYVGMRPTMH
                     VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEVNRGAKRVLQSVVIEVAGEEQVTF
                     DKFEAAQLSGLDREVIKTQLVDSGLWTALRTRPFSKVPAIESSTKAIFVTAMDTNPLA
                     AKPELIINEQQEAFIAGLDILSALTEGKVYVCKSGTSLPRSSQSNVEEHVFDGPHPAG
                     LAGTHMHFLYPVNAENVAWSINYQDVIAFGKLFLTGELYTDRVVSLAGPVVNNPRLVR
                     TVIGASLDDLTDNELMPGEVRVISGSVLTGTHATGPHAYLGRYHQQVSVLREGREKEL
                     FGWAMPGKNKFSVTRSFLGHVFKGQLFNMTTTTNGSDRSMVPIGNYERVMPLDMEPTL
                     LLRDLCAGDTDSAQALGALELDEEDLALCTFVCPGKYEYGTLLRECLDTIEKEG"
conflict             1205
                     /gene="nqr1"
                     /citation=[2]
                     /replace="c"
conflict             1234
                     /gene="nqr1"
                     /citation=[2]
                     /replace="g"
conflict             1256
                     /gene="nqr1"
                     /citation=[2]
                     /replace="t"
gene                 1732. .2976
                     /gene="nqr2"
CDS                  1732. .2976
                     /gene="nqr2"
                     /note="hydrophobic membrane protein with pI 8.14; nqrB,
                     Nqr2 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr2 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22911.1"
                     /db_xref="GI:2558474"
                     /translation="MALKKFLEDIEHHFEPGGKHEKWFALYEAVATVFYTPGIVTNKS
                     SHVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGDQLATVISGNWHY
                     WLTEMLGGTIAADAGVGSKMLLGATYFLPIYATVFLVGGFWEVLFCMVRKHEVNEGFF
                     VTSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPALAGRAFLFFAYPA
                     QISGDVVWTAADGFSGATALSQWAQGGNGALVNTVTGSPITWMDAFIGNIPGSIGEVS
                     TLALMIGAAMIVYMRIASWRIIAGVMIGMIAVSTLFNVIGSDTNPMFNMPWHWHLVLG
                     GFAFGMFFMATDPVSASFTNKGKWWYGILIGAMCVMIRVVNPAYPEGMMLAILFANLF
                     APLFDHVVIEKNIKRRLARYGK"
conflict             1881
                     /gene="nqr2"
                     /citation=[2]
                     /replace="c"
conflict             2400
                     /gene="nqr2"
                     /citation=[2]
                     /replace="t"
conflict             2417. .2418
                     /gene="nqr2"
                     /citation=[2]
                     /replace="tc"
conflict             2614
                     /gene="nqr2"
                     /citation=[2]
                     /replace="g"
gene                 2966. .3736
                     /gene="nqr3"
CDS                  2966. .3736
                     /gene="nqr3"
                     /note="acidic protein with pI 4.83; nqrC, Nqr3 subunit of
                     Na-NQR complex gamma-subunit"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr3 subunit of Na-translocating NADH-quinone
                     reductase complex gamma-subunit"
                     /protein_id="BAA22912.1"
                     /db_xref="GI:2558475"
                     /translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDKQKANAVLD
                     KQSKIVEVAGIDANGKKVPELFAEYIEPRLVDLETGNFTEGNASTYDQREASKDAERS
                     IALTPEEDVADIRRRANTAVVYLVKDQDEVQKVILPMHGKGLWSMMYAFVAVETDGNT
                     VSAITYYEQGETPGLGGEVENPSWRDQFIGKKLYNEDHQPAIKVVKGGAPQGSEHGVD
                     GLSGATLTSNGVQHTFDFWLGDKGFGPFLAKVRDGELN"
conflict             3295
```

```
                     /gene="nqr3"
                     /citation=[2]
                     /replace="c"
     conflict        3514. .3515
                     /gene="nqr3"
                     /citation=[2]
                     /replace="ta"
     gene            3736. .4368
                     /gene="nqr4"
     CDS             3736. .4368
                     /gene="nqr4"
                     /note="hydrophobic membrane protein with pI 9.17; nqrD,
                     Nqr4 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr4 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22913.1"
                     /db_xref="GI:2558476"
                     /translation="MSSAQNVKKSILAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                     FGLEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
     gene            4376. .4972
                     /gene="nqr5"
     CDS             4376. .4972
                     /gene="nqr5"
                     /note="hydrophobic membrane protein with pI 7.00; nqrE,
                     Nqr5 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr5 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22914.1"
                     /db_xref="GI:2558477"
                     /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGVA
                     VVVVLTIAVPVNNLVYNLVLRENALVEGVDLSFLNFITFIGVIAALVQILEMVLDRFF
                     PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYNFAESIVYGFGSGVGWMLAIVALAGI
                     REKMKYSDVPPGLRGLGITFITVGLMALGFMSFSGVQL"
     conflict        4410
                     /gene="nqr5"
                     /citation=[2]
                     /replace=""
     conflict        4419
                     /gene="nqr5"
```

```
  Query Match             35.2%;  Score 473.6;  DB 1;  Length 6908;
  Best Local Similarity   60.0%;  Pred. No. 2e-105;
  Matches  807;  Conservative    0;  Mismatches  534;  Indels    3;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          ||||||| ||| || || ||  |  | || || ||||| || |  ||    || |  ||
Db    386 ATGATTACAATAAAGAAGGGCTTGGATCTTCCTATCGCAGGAACTCCATCCCAGGTGATT 445

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            | || ||   | |||| |   ||||||| ||||||||||||||| || |  || |||||
Db    446 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 505

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||  | ||| | |||   |  || ||||  || |||||||  ||||| || ||||||||
Db    506 CCTACCATGCATGTCCGCGTTGGTGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAT 565

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          |||||||| || ||||||   |||||||| || ||  | || ||| |   ||   || ||
Db    566 AAAAAGAACCCTGGCGTGAAATTTACTGCACCAGCAGCCGGTAAAGTGATCGAAGTTAAC 625

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          |||||||  || || || ||||| || || |||||||  || | |||  | ||  |  |
Db    626 CGTGGCGCTAAACGTGTCCTTCAATCTGTAGTGATTGAAGTGGCAGGTGAAGAGCAGGTG 685

Qy    301 gagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgc 360
```

```
Db    686 ACATTTGATAAGTTCGAAGCCGCTCAACTTTCAGGTCTAGATCGTGAAGTGATCAAGACT 745

Qy    361 aacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcct 420
            |  || |  | || ||  | ||||| ||  | || || |||||||| |||||  |||||
Db    746 CAGTTGGTTGACTCTGGCCTATGGACCGCTTTACGTACTCGTCCGTTTAGCAAGGTTCCT 805

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||  |||| |     |    || || ||||| | ||| ||||| || |||||  | ||
Db    806 GCAATCGAGTCTTCTACTAAGGCTATTTTCGTAACTGCAATGGATACTAATCCATTAGCA 865

Qy    481 gccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggta 540
          ||  | |||  | | || || || ||       || |  ||||    || ||     |
Db    866 GCTAAGCCTGAGTTGATCATTAATGAGCAACAAGAAGCATTCATTGCTGGTCTAGACATT 925

Qy    541 ttgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccg 600
            |        |||| ||| ||||  |  | || ||||  | ||||| |     ||||
Db    926 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAA---ATCTGGCACTAGTTTGCCA 982

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttg 660
                  | ||| | |||  ||| |  ||||  ||||| || || || ||| |
Db    983 CGCTCTTCACAGTCTAACGTAGAAGAACACGTCTTCGATGGCCCTCACCCAGCAGGTCTT 1042

Qy    661 agtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatc 720
            |||||| ||||| |||||| |  | || ||    ||  | ||    |  |||| ||||
Db   1043 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCATGGAGCATC 1102

Qy    721 aattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgag 780
          || || ||||| || ||  |  | ||     ||||   ||| ||    ||  |||| ||
Db   1103 AACTACCAAGACGTTATCGCGTTCGGTAAGCTGTTCCTAACTGGTGAGCTATACACAGAT 1162

Qy    781 cgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          || ||  |  | |||| |||| |    || || || || ||  |  | ||||||||  |
Db   1163 CGTGTTGTGTCTCTAGCTGGTCCAGTTGTGAATAACCCTCGTTTAGTTCGTACCGTAATC 1222

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagacaaccgcgtgatt 900
          ||||| |   ||    |  | || |   ||| || |       ||    || ||||||
Db   1223 GGTGCAAGCTTAGATGACCTAACAGATAACGAGCTGATGCCTGGCGAAGTTCGTGTGATT 1282

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          || ||||| ||| |||  ||  |   | ||   ||  | ||||  ||  | || || |||
Db   1283 TCTGGTTCTGTACTGACAGGTACACATGCAACTGGTCCACACGCATACCTTGGCCGTTAC 1342

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||  | ||| |||| ||  |    |||||||     ||||| || || |||||||    |
Db   1343 CATCAACAGGTTTCTGTATTACGTGAAGGCCGTGAGAAAGAACTATTTGGCTGGGCAATG 1402

Qy   1021 ccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaa 1080
          ||      | | || | |||  | || ||   |   ||| || ||  |  | |||    |
Db   1403 CCTGGTAAGAATAAGTTCTCGGTAACTCGCTCTTTCCTTGGTCACGTATTCAAAGGTCAG 1462

Qy   1081 ctcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtact 1140
            | |||||  | |  ||  |   ||| ||  | || ||  | ||||| || ||||| |
Db   1463 TTGTTCAATATGACAACGACGACCAATGGTAGTGATCGTTCAATGGTTCCAATTGGCAAC 1522

Qy   1141 tacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtc 1200
          ||||| ||||| |||||  | |||||   ||| || |||||  | || ||| ||   |
Db   1523 TACGAACGCGTAATGCCTCTAGATATGGAGCCTACTCTGCTGCTTCGTGATCTATGTGCA 1582

Qy   1201 ggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgct 1260
          || ||||| ||||| || || ||| |||||     | ||| |||| |||||||| || ||
Db   1583 GGTGATACAGACAGTGCTCAAGCACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 1642

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          ||||| | ||| || || || || || ||||| |||||  || |  | ||  ||   ||
Db   1643 TTGTGTACCTTTGTATGTCCAGGTAAGTACGAGTACGGTACGTTACTTCGTGAATGCCTA 1702

Qy   1321 gaaaccattgagaaggaaggctga 1344
          || || || ||||||||||| | |
Db   1703 GATACTATCGAGAAGGAAGGGTAA 1726
```

RESULT 17

```
AX067460/c
LOCUS       AX067460               96109 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 35 from Patent WO0078968.
ACCESSION   AX067460
VERSION     AX067460.1  GI:12545080
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 96109)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 35 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .96109
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
BASE COUNT    28783 a  18910 c  20341 g  28075 t
ORIGIN

  Query Match           24.3%;  Score 326.4;  DB 6;  Length 96109;
  Best Local Similarity  54.2%;  Pred. No. 2.9e-69;
  Matches  735;  Conservative    0;  Mismatches  606;  Indels   15;  Gaps    3;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          |||||||  ||||||||||  |  | ||||| ||| | |||   ||     |    ||
Db  85201 ATGATTACCATCAAAAAAGGCTTGGATCTGCCAATCACTGGCGAGCCCTCCAAGCAGATT 85142

Qy     61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            |||   | || ||| ||   | || |||||  | ||| | || |||| ||| |||||
Db  85141 AGCGAGCACACGCCCACTA---AGGTGGCGTTAGTCGGCTACGATTATGTCGGCATGCGT 85085

Qy    121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||  | ||| | ||||| ||||| |||   ||||  ||||| || ||||| |  ||||||
Db  85084 CCGACCATGCATGTCAAAGAAGGTGATAAGGTCACCAAAGGACAGGTGCTTTACGAAGAC 85025

Qy    181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcac 240
          |||||     | |||||  | | ||| || || | |   ||||   |||  ||  | |
Db  85024 AAAAAACGAGCAGGCGTCCTTTATACAGCACCAGTTGGCGGCACTGTCGTTGCTGTCAAT 84965

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttg------aaggcaacgac 294
          |||||||||   || ||| || |     | || || ||||| |      | |  | .||
Db  84964 CGTGGCGAACGACGAGTATTTGAAAGCTTGGTCATCGCCGTCGATCCAAATGCAAGTGAA 84905

Qy    295 gaaatcgagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtg 354
          ||||||   ||| |    ||     ||||||   || |||   |     || |  |||
Db  84904 GAAATCACTTTTAATGCTTATCGCTCTGAAGATTTGTCAATGCTGGATCGCAAGACAGTT 84845

Qy    355 cgccgcaacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaa 414
              |  || ||   || |||   ||||| || || ||||| ||||| || |||
Db  84844 GTTGAGCAGTTGGTCGCTTCAGGTGAATGGACCGCCCTACGCACTCGTCCTTTTAGCCGT 84785

Qy    415 attcctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccg 474
          |  || |   |  ||  |   || |  ||||| || || |  || |  || ||||||||
Db  84784 ACACCAGAAATTAATAGCATACCTTCTGCCATTTTTGTGACGGCAACAGATACCAATCCT 84725

Qy    475 ctggctgccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctg 534
           ||||||  ||||||  |   | |  |    ||||      | |  || ||    ||| ||
Db  84724 TTGGCTGTTGACCCTGCCACCGTATTATTTGAAGAAATACAAGCATTTAATGATGGCTTG 84665

Qy    535 ttggtattgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagac 594
            | ||       |  ||  |   |          |  |   |    ||   | ||
Db  84664 GCAATTTTATCAACGCTATCCCCAAAAACCTATGTCTGTCATGGTGATATTGTACCTGAG 84605

Qy    595 gtgccgtctgaaaatgctgccaacatcgaaacac------atgaattcggcggcccgcat 648
              | |||  |||| |   | ||| ||         | |  ||    ||    |||
Db  84604 ACCTTAAGTAAAACCGCTGACGGTAACGATACCATCTATGAAGGGTTTACTGGTGTACAT 84545

Qy    649 cctgccggtttgagtggcacgcacattcatttcatcgagccggtcggcgcgaataaaacc 708
          || || ||       || ||||||||||| ||  |  | ||  | |        |

Db  84544 CCAGCAGGCAACGCAGGTACGCACATTCACTTTTTACACCCTTTGGCACGAGGCAGCTTA 84485

Qy    709 gtgtggaccatcaattatcaagatgtaattaccattggccgtttgtttgcaacaggccgt 768
          || ||||| || |||||||||||||| ||  |||||||     | ||| | ||||| ||
Db  84484 GTTTGGACGATTAATTATCAAGATGTCATCGCCATTGGTAAGCTATTTACCACAGGTCGC 84425

Qy    769 ctgaacaccgagcgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttg 828
            |  |||| |||||  ||||     | |  ||| |   |||   ||| || ||| |  ||
Db  84424 ATTTACACAGAGCGTCTGATCAGTTTGGCAGGTCCAGCAGTGGCCAATCCTCGCTTGGTG 84365

Qy    829 cgtaccgttttgggtgcgaaagtatcgcaaattactgcgggcgaattggttgacacagac 888
          ||||||  |   |||||  |  |  |  || | ||    || ||| |    | |   ||
Db  84364 CGTACCACTCGAGGTGCTGATTTGACCGAATTGACCCAAGGTGAACTCAACGGCGACGAT 84305

Qy    889 aaccgcgtgatttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattat 948
          |||||  || ||||| || || ||  |    ||    |   |    || ||    | | |
Db  84304 AACCGTGTCATTTCAGGCTCTGTTCTATCAGGTCGCACAGCTGTTGGTGCTACGGCATTT 84245

Qy    949 ttgggacgctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttc 1008
          |||||  |||  || |||||  | || ||  |    |||||  ||              |
Db  84244 TTGGGCCGCTTTCATAATCAAGTCTCTGTACTTACCGAAGGTCGTGAACGCCCTGCACTT 84185

Qy   1009 ggctgggttgcgccgcagccggacaaatactccatcacgcgtacaaccctcggccatttc 1068
            ||   || | ||   | |  |    |  || |    ||  |       |    || ||
Db  84184 CACTTTTTTACACCAGGGGCAAATCGTTTTTCAAAATTGCCAATTTATATTTCACAGTTT 84125

Qy   1069 ctgaaaaacaaactcttcaagttcaacacagccgtcaacggcggcgaccgcgccatggtg 1128
            | |||  ||||   | | | || |  ||      |||||||     || ||||||||
Db  84124 TTTAAAGGCAAAAAATACGATTTTACGACCAGTACCAACGGCTCGCCTCGTGCCATGGTA 84065

Qy   1129 ccgattggtacttacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgcgc 1188
          || || ||    | ||   || |||||    ||    || || |||| |||  ||||
Db  84064 CCAATCGGCTTATTTGAAACTGTCATGCCTCAAGACTATCTACCTACCCAGCTACTGCGT 84005

Qy   1189 gatttaatcgtcggcgataccgacagcgcgcaggcattgggttgcttggaattggacgaa 1248
          | ||| || |||| |||| |  ||| |||   | |||||||    | ||  | || |||
Db  84004 GCTTTGATTGTCGAAGATATCATCAGTGCGGTTGAATTGGGTGCACTTGAGCTTGATGAA 83945

Qy   1249 gaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttg 1308
          |||||  | ||  |||| |  ||||| |  || |||||||| || | |||     | |||
Db  83944 GAAGATTTGGCACTGTGTACTTTCGTATCGCCAGGCAAATATGAGTTCGGTGATATTTTG 83885

Qy   1309 cgcaaagtgctggaaaccattgagaaggaaggctga 1344
          ||  |    ||      |||||| |||||||| |
Db  83884 CGTGATAACTTGACCCGTATTGAGCAGGAAGGCTAA 83849


RESULT  18
AF144076
LOCUS       AF144076               1356 bp    DNA     linear   BCT 01-JUN-1999
DEFINITION  Porphyromonas gingivalis strain W50 50 kD antigen PG1 gene,
            complete cds.
ACCESSION   AF144076
VERSION     AF144076.1  GI:4929279
KEYWORDS    .
SOURCE      Porphyromonas gingivalis.
  ORGANISM  Porphyromonas gingivalis
            Bacteria; CFB group; Bacteroidetes; Bacteroidales;
            Porphyromonadaceae; Porphyromonas.
REFERENCE   1  (bases 1 to 1356)
  AUTHORS   Ross,B.C., Barr,I., Patterson,M., Agius,C., Rothel,L., Margetts,M.,
            Hocking,D. and Webb,E.
  TITLE     Porphyromonas gingivalis polypeptides and nucleic acids
  JOURNAL   Patent: Australia (AU 98/01023)-PCT 10-DEC-1998;
            CSL Ltd;
            45 Poplar Road;
            R&D;
            Parkville, Victoria;
            Australia;
REFERENCE   2  (bases 1 to 1356)
  AUTHORS   Ross,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1999) R&D, CSL Ltd, 45 Poplar Road, Parkville,
```

```
          Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .1356
                     /organism="Porphyromonas gingivalis"
                     /strain="W50"
                     /db_xref="taxon:837"
     CDS             1. .1356
                     /codon_start=1
                     /transl_table=11
                     /product="50 kD antigen PG1"
                     /protein_id="AAD33930.1"
                     /db_xref="GI:4929280"
                     /translation="MAKVIKTKKGLALNLKGKPLPEMLAEPAQSPTYAVVPDDFEGVI
                     PKVTARPGDKVRAGSALMHHKAYPEMKFTSPVSGEVIAVNRGAKRKVLSIEVKPDGLN
                     EYESFPVGDPSALSAEQIKELLLSSGMWGFIKQRPYDIVATPDIAPRDIYITANFTAP
                     LAPDFDFIVRGEERALQTAIDALAKLTTGKVYVGLKPGSSLGLHNAEIVEVHGPHPAG
                     NVGVLINHTKPINRGETVWTLKATDLIVIGRFLLTGKADFTRMLAMTGSDAAAHGYVR
                     IMPGCNVFASFPGRLTIKESHERVIDGNVLTGKKLCEKEPFLSARCDQITVIPEGDDV
                     DELFGWAAPRLDQYSMSRAYFSWLQGKNKEYVLDARIKGGERAMIMSNEYDRVFPMDI
                     YPEYLLKAIIAFDIDKMEDLGIYEVAPEDFATCEFVDTSKIELQRIVREGLDMLYKEM
                     N"
BASE COUNT      338 a    333 c    380 g    305 t
ORIGIN
```

Query Match 8.5%; Score 113.6; DB 1; Length 1356;
Best Local Similarity 46.1%; Pred. No. 3.6e-17;
Matches 623; Conservative 0; Mismatches 699; Indels 30; Gaps 6;

```
Qy       2 tgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgttt 61
Db      11 TTATAAAAACAAAAAAAGGCCTTGCACTTAATCTGAAAGGAAAACCGCTGCCCGAGATGC 70

Qy      62 acgacggcccggcc---attaccgaagtcgcgttgcttggcgaagaatatgccggtatgc 118
Db      71 TGGCCGAACCGGCCCAAAGTCCTACTTACGCGGTCGTGCCCGACGATTTTGAAGGTGTTA 130

Qy     119 gcccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaag 178
Db     131 TCCCCAAGGTGACGGCTCGTCCGGGGGATAAGGTGCGTGCCGGCTCAGCACTGATGCACC 190

Qy     179 acaaaaagaatccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattc 238
Db     191 ACAAGGCATATCCGGAGATGAAGTTTACAAGTCCGGTTAGCGGCGAAGTGATCGCGGTGA 250

Qy     239 accgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaa 298
Db     251 ATCGCGGTGCCAAGCGCAAGGTGTTGAGCATCGAGGTGAAACCGGACGGACTGAACGAAT 310

Qy     299 tcgagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgcc 358
Db     311 ACGAGT------CATTCCCTGTCGGGGATCCGTCTGCCCTCTCTGCCGAACAGATCAAGG 364

Qy     359 gcaacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattc 418
Db     365 AGCTTTTACTGTCGAGCGGTATGTGGGGTTTTATTAAGCAACGTCCTTACGACATAGTGG 424

Qy     419 ctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctgg 478
Db     425 CTACACCGGATATAGCTCCACGCGACATTTATATTACTGCCAACTTTACTGCACCATTGG 484

Qy     479 ctgccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttgg 538
Db     485 CTCCGGACTTCGATTTCATCGTTCGAGGAGAAGAACGCGCCCTGCAGACTGCCATCGATG 544

Qy     539 tattgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgc 598
Db     545 CCTTGGCCAAACTCACGACAGGAAAGGTGTATGTGGGCCTGAAGCCGGGTTCATCTCTGG 604

Qy     599 cgtctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtt 658
Db     605 GCTTGCACAATGCAGAAATCGTAGAAGTACA---------CGGACCTCATCCGGCAGGTA 655

Qy     659 tgagtggcacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggacca 718
Db     656 ACGTGGGCGTGCTGATCAATCATACGAAGCCAATCAATCGGGGCGAAACGGTGTGGACGC 715

Qy     719 tcaattatcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccg 778
Db     716 TCAAGGCTACCGACCTGATCGTGATCGGACGTTTCCTGCTTACGGGCAAAGCCGATTTTA 775

Qy     779 agcgcgtgattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttt 838
Db     776 CCAGAATGATTGCCATGACCGGCTCAGACGCTGCAGCTCACGGATACGTCCGTATTATGC 835

Qy     839 tgggtgcgaaagtatcgcaaattactgcgggcgaattg------gttgacacagacaacc 892
Db     836 CGGGTTGCAATGTCTTTGCTTCCTTCCCCGGCCGACTGACAATAAAGGAATCTCACGAGC 895

Qy     893 gcgtgatttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgg 952
Db     896 GTGTGATCGATGGCAATGTGCTGACCGGTAAGAAGCTCTGCGAGAAGGAGCCTTTCCTGT 955

Qy     953 gacgctaccacaatcagatttccgttatcgaagaaggccgc---agcaaagagctgttcg 1009
Db     956 CAGCCCGGTGTGACCAGATCACGGTGATCCCCGAAGGCGACGATGTGGACGAACTCTTCG 1015

Qy    1010 gctgggttgcgccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcc 1069
Db    1016 GGTGGGCTGCACCCCGTCTCGATCAGTACAGCATGAGCAGAGCTTATTTCTCTTGGTTGC 1075

Qy    1070 ---tgaaaaacaaactcttcaagttcaacacagccgtcaacggcggcgaccgcgccatgg 1126
Db    1076 AGGGGAAAAACAAAGAGTACGTACTCGATGCCCGGATCAAGGGTGGCGAACGTGCTATGA 1135

Qy    1127 tgccgattggtacttacgagcgcgtgatgcccttggatatcctgcccaccctgcttttgc 1186
Db    1136 TCATGAGCAACGAGTATGACCGCGTTTTCCCGATGGACATCTATCCGGAGTATTTGCTCA 1195

Qy    1187 gcgatttaatcgtcggcgataccgacagcgcgcaggcattgggttgcttggaattggacg 1246
Db    1196 AGGCTATTATAGCATTCGACATCGACAAGATGGAGGACTTAGGCATATATGAAGTGGCTC 1255

Qy    1247 aagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgt 1306
Db    1256 CGGAGGACTTTGCCACTTGCGAATTTGTGGATACATCCAAGATCGAGCTGCAGCGTATCG 1315

Qy    1307 tgcgcaaagtgctggaaaccattgagaaggaa 1338
Db    1316 TTCGCGAGGGCTTGGATATGCTCTATAAGGAA 1347
```

```
RESULT 19
HIU20229/c
LOCUS       HIU20229                4926 bp    DNA     linear   BCT 09-FEB-1995
DEFINITION  Haemophilus influenzae BOLA (bolA), glutathione reductase (gor),
            phosphatidylserine decarboxylase (psd), 30K protein (rpmF), genes,
            complete cds.
ACCESSION   U20229
VERSION     U20229.1  GI:644850
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 4926)
  AUTHORS   Barcak,G.J. and Heimer,S.R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4926)
  AUTHORS   Barcak,G.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-1995) Gerard J. Barcak, Biochemistry, University
            of Maryland at Baltimore, 108 North Greene Street, Baltimore, MD
            21201, USA
FEATURES             Location/Qualifiers
     source          1. .4926
```

```
                     /organism="Haemophilus influenzae"
                     /strain="Rd"
                     /isolate="KW20"
                     /db_xref="taxon:727"
     CDS             complement(27. .224)
                     /note="orf65"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62134.1"
                     /db_xref="GI:644851"
                     /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                     VREGDVVKKGQVLFEDKKILV"
     gene            complement(519. .842)
                     /gene="bolA"
     CDS             complement(519. .830)
                     /gene="bolA"
                     /function="stationary phase regulatory protein"
                     /codon_start=1
                     /transl_table=11
                     /product="BOLA"
                     /protein_id="AAA62135.1"
                     /db_xref="GI:644852"
                     /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                     SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                     R"
     misc_feature    775. .785
                     /note="encodes DNA uptake recognition sequence"
     RBS             complement(838. .842)
                     /gene="bolA"
     protein_bind    877. .913
                     /note="OxyR recongition sequence (Tartaglia, L.A., et al.,
                     J. Mol Biol. 210: 709-710, 1989)"
                     /bound_moiety="OxyR"
     RBS             915. .919
                     /note="orf199"
     CDS             924. .1523
                     /note="orf199"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62136.1"
                     /db_xref="GI:644853"
                     /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                     VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                     QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                     VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
     gene            1343. .3109
                     /gene="gor"
     -35_signal      1343. .1348
                     /gene="gor"
                     /note="putative"
     -10_signal      1365. .1370
                     /gene="gor"
                     /note="putative"
     -35_signal      1585. .1590
                     /gene="gor"
                     /note="putative"
     -10_signal      1605. .1610
                     /gene="gor"
                     /note="putative"
     RBS             1728. .1731
                     /gene="gor"
     CDS             1739. .3109
                     /gene="gor"
                     /EC_number="1.6.4.2"
                     /function="reduces oxidized glutathione using NADPH"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="glutathione reductase"
                     /protein_id="AAA62137.1"
                     /db_xref="GI:644854"
                     /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                     VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                     LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                     DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                     LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                     LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                     KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                     QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                     EFVTMR"
     gene            complement(3218. .4090)
                     /gene="psd"
     CDS             complement(3218. .4090)
                     /gene="psd"
                     /function="lipid metabolism"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphatidylserine decarboxylase"
                     /protein_id="AAA62138.1"
                     /db_xref="GI:644855"
                     /translation="MMTSNSYWQRLKVAFQYVMPQIYLTQIAGWFAKQKWGKITHFVI
                     KAFAKKYNIDMSIAQKEQFSDYASFNEFFIRPLKENARPINQNPTALCLPADGRISEC
                     GHIDDNLLLQAKGHFFSLEDLLAEDKELVETFKNGEFVTTYLSPRDYHRVHMPCDATL
                     RKMIYVPGDLFSVNPFLAQHVPNLFARNERVICVFDTEFGTMVQILVGATITASIGTT
                     WAGVINPPRHNEVKTWTYEGESAVKLLKGQEMGWFQLGSTVINLFQANQVRLADHLSV
                     NEPVRMGEILAYKK"
     misc_feature    complement(3364. .3374)
                     /gene="psd"
                     /note="encodes DNA uptake recognition sequence"
     misc_feature    complement(3780. .3790)
                     /gene="psd"
                     /note="encodes DNA uptake recognition sequence"
     misc_feature    3818. .3825
                     /note="encodes DNA uptake recognition sequence"
     RBS             complement(4097. .4100)
                     /note="psd"
     protein_bind    complement(4126. .4152)
                     /note="OxyR recognition sequence (Toledano, M.B. et al.,
                     Cell 78:897-909.)"
                     /bound_moiety="OxyR"
     gene            4247. .4771
                     /gene="rpmF"
     CDS             4247. .4771
                     /gene="rpmF"
                     /codon_start=1
                     /transl_table=11
                     /product="30K protein"
                     /protein_id="AAA62139.1"
                     /db_xref="GI:644856"
                     /translation="MQKVKLPLTVDPIKDAQRRLDYVGYYAANQLERLAESVVNVLSD
                     AQVTLSFYVDPQKLVVMKGKVQIDVDLECQRCNEPYKQTLECEFTYSPVANWDQADDL
                     PEIYEPIEFNEFGEIDLIGTVEDELILALPLVPMHSSEHCEVSAQEQVFGVLPEELAK
                     KPNPFAVLANLKQK"
     CDS             complement(4561. .>4926)
                     /note="orf121"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62140.1"
                     /db_xref="GI:644857"
                     /translation="LQLRGGVSEFHRLLYLLIQQQLLKHRVNDAYHDVNEIFCFVEQP
                     LAILLNLIYFCFKLANTANGFGFFANSSGNTPKTCSCADTSQCSDECIGTSGKAKISS
                     SSTVPIKSISPNSLNSIGS"
BASE COUNT     1598 a    862 c    938 g   1528 t
ORIGIN


  Query Match             7.5%;  Score 100.6;  DB 1;  Length 4926;
  Best Local Similarity   68.6%;  Pred. No. 5.8e-14;
  Matches  153;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagccgtt 60
          ||||||| ||| || |||||| |  | || || || ||||| | ||| |  ||||   |
Db    224 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 165
```

```
Qy      61 tacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           |   ||||   ||  | |   |||| ||| | || || || || ||||  || ||||||
Db     164 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 105

Qy     121 ccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           || || ||||| ||    ||||||||||  || ||||||| ||||| || ||||||||
Db     104 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTGAAGAC 45

Qy     181 aaaaagaatccgggcgtggtgtttactgcgccggcttcaggca 223
           ||||| ||||| || ||  | ||||| || || ||    || |
Db      44 AAAAA-AATCCTGGTGTAATTTTTACAGCCCCTGCAAGCGGTA 3
```

```
RESULT 20
AF429315
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN
```

```
  Query Match             3.5%;  Score 47.4;  DB 9;  Length 125020;
  Best Local Similarity   11.4%;  Pred. No. 0.76;
  Matches   45;  Conservative  172;  Mismatches  177;  Indels    0;  Gaps    0;

Qy     239 accgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaa 298
           ::: ::::: : : |: :  ::   : | | :|:|   :::| :  :::   ::  :: :
Db   17040 RSYWKRSMKCWSRMCSRSRSSKCKSRGCGGSMGKGKGGKSYGRKTKRKSRGKMKGAKWYM 17099

Qy     299 tcgagtttgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgcc 358
           :::::  :  ::::: : |::|: : : ||: :|:     :  :  :   : : ::| :
Db   17100 YYRRSMRKMKMYSSKGMYCMYCWCWGRRGCYCSCMTSRSAMCCSYCYAKCKSMCYSCYYG 17159

Qy     359 gcaacctgatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattc 418
           ::  :: | :::  ::::::  : :  :|:|||       ||  |   |  | ||  ||
Db   17160 MSMKGYYYACSYRGSMSSKYCMRGSTYSTSCGCCCTTTTTCCCCCNANTGGGGAAGCTTT 17219

Qy     419 ctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctgg 478
           :|:: ::      : ::  ::  :: ::  |  :|  ::: : :::| :| | :
Db   17220 TNCNKKTYSYYRKRNGCAMCKYNNYYNNSWRSSCRAGMSCTKYKSSMTMSMASYCWCMSM 17279

Qy     479 ctgccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttgg 538
           :: ::: : :  ::  :  ::    :::  :::||   :: : ::: ::::|:::  : :
Db   17280 YYCSMRSMASRGMSWSSYMKMMKSSWRMSCYMKCCWCMKMCYCMMRSMRSRGMSYYMYAS 17339

Qy     539 tattgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgc 598
           :   ::||: |   :|::  : :     ::: :  :|  :::|:|:  |: ::: | : :
Db   17340 WKSSSRGCYCTRCYWCMSSKSCYKSYYMMMRSKRMKGMKWSRCWGSMWGSASRSSCYKCY 17399

Qy     599 cgtctgaaaatgctgccaacatcgaaacacatga 632
           :  :::  :  :    :::: | ::::  :::
Db   17400 KSMRCSMMSSKCYRCAGCMMMKGGYMRYMRCWSM 17433
```

```
RESULT 21
BACGNTZA/c
LOCUS       BACGNTZA              36200 bp    DNA     linear   BCT 05-FEB-1999
DEFINITION  Bacillus subtilis 36kb sequence between gntZ and trnY genes
            encoding 34 ORFs.
ACCESSION   D78193
VERSION     D78193.1  GI:1064780
KEYWORDS    phospho-bete-glucosidase; 6-phosphogluconate dehydrogenase; alkyl
            hydroperoxide reductase; amino acid permease; arginase; orthinine
            aminotransferase; positive regulatory protein; gntZ; ahp; bglA;
            yyd; yyc; formaldehyde dehydrogenase.
SOURCE      Bacillus subtilis (strain:168) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 36200)
  AUTHORS   Ogasawara,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1995) Naotake Ogasawara, Nara Institute of
            Science and Technology, Graduate School of Biological Sciences;
            8916-5 Takayama-cho, Ikoma, Nara 630-01, Japan
            (E-mail:nogasawa@bs.aist-nara.ac.jp, Tel:07437-2-5430,
            Fax:07437-2-5439)
REFERENCE   2  (sites)
  AUTHORS   Zhang,J. and Aronson,A.
  TITLE     A Bacillus subtilis bglA gene encoding phospho-beta-glucosidase is
            inducible and closely linked to a NADH dehydrogenase-encoding gene
  JOURNAL   Gene 140 (1), 85-90 (1994)
  MEDLINE   94171085
REFERENCE   3  (sites)
  AUTHORS   Hartford,O.M. and Dowds,B.C.
  TITLE     Isolation and characterization of a hydrogen peroxide resistant
            mutant of Bacillus subtilis
  JOURNAL   Microbiology 140 (Pt 2), 297-304 (1994)
  MEDLINE   94236234
REFERENCE   4  (sites)
  AUTHORS   Ogasawara,N., Nakai,S. and Yoshikawa,H.
  TITLE     Systematic sequencing of the 180 kilobase region of the Bacillus
            subtilis chromosome containing the replication origin
  JOURNAL   DNA Res. 1 (1), 1-14 (1994)
  MEDLINE   96051385
REFERENCE   5  (sites)
```

```
  AUTHORS   Calogero,S., Gardan,R., Glaser,P., Schweizer,J., Rapoport,G. and
            Debarbouille,M.
  TITLE     RocR, a novel regulatory protein controlling arginine utilization
            in Bacillus subtilis, belongs to the NtrC/NifA family of
            transcriptional activators
  JOURNAL   J. Bacteriol. 176 (5), 1234-1241 (1994)
  MEDLINE   94156824
REFERENCE   6  (sites)
  AUTHORS   Yoshida,K., Seki,S., Fujimura,M., Miwa,Y. and Fujita,Y.
  TITLE     Cloning and sequencing of a 36-kb region of the Bacillus subtilis
            genome between the gnt and iol operons
  JOURNAL   DNA Res. 2 (2), 61-69 (1995)
  MEDLINE   96093926
REFERENCE   7  (sites)
  AUTHORS   Gardan,R., Rapoport,G. and Debarbouille,M.
  TITLE     Expression of the rocDEF operon involved in arginine catabolism in
            Bacillus subtilis
  JOURNAL   J. Mol. Biol. 249 (5), 843-856 (1995)
  MEDLINE   95311309
REFERENCE   8  (bases 1 to 36200)
  AUTHORS   Kasahara,Y., Nakai,S., Yoshikawa,H. and Ogasawara,N.
  TITLE     36kb sequence between gntZ and trnY of B. subtilis genome
  JOURNAL   Unpublished (1997)
FEATURES             Location/Qualifiers
     source          1. .36200
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            1. .32
                     /gene="gntZ"
     CDS             <1. .32
                     /gene="gntZ"
                     /citation=[6]
                     /codon_start=3
                     /transl_table=11
                     /product="6-phosphogluconate dehydrogenase"
                     /protein_id="BAA11267.1"
                     /db_xref="GI:1064781"
                     /translation="GVFHTNWSE"
     gene            496. .1059
                     /gene="ahpC"
     CDS             496. .1059
                     /gene="ahpC"
                     /note="induced by hydrogen peroxide"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="alkyl hydroperoxide reductase"
                     /protein_id="BAA11268.1"
                     /db_xref="GI:1064782"
                     /translation="MSLIGKEVLPFEAKAFKNGEFIDVTNEDLKGQWSVFCFYPADFS
                     FVCPTELEDLQEQYAALKELGVEVYSVSTDTHFVHKGWHDSSEKISKITYAMIGDPSQ
                     TISRNFDVLDEETGLADRGTFIIDPDGVIQTVEINAGGIGRDASNLVNKVKAAQYVRQ
                     NPGEVCPAKWEEGGETLTPSLDLVGKI"
     gene            1073. .2602
                     /gene="ahpF"
     CDS             1073. .2602
                     /gene="ahpF"
                     /note="induced by hydrogen peroxide"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="alkyl hydroperoxide reductase"
                     /protein_id="BAA11269.1"
                     /db_xref="GI:1064783"
                     /translation="MVLDANIKAQLNQYMQLIENDIVLKVSAGEDDTSKDMLALVDEL
                     ASMSSKISVEKAELNRTPSFSVNRVGEDTGVTFAGIPLGHEFTSLVLALLQVSGRPPK
                     VDQKVIDQVKKISGEYHFESYISLTCHNCPDVVQALNMMSVLNPNITHTMIDGAAYKA
                     EVESKNIMAVPTVYLNGESFGSGRMTLEEILAKMGSGTDASEFADKEPFDVLVVGGGP
                     AGASAAIYTARKGIRTGVVAERFGGQVLDTMSIENFISVKATEGPKLAASLEEHVKEY
                     DIDVMNLQRAKRLEKKDLFELELENGAVLKSKTVILSTGARWRNVNVPGEQEFKNKGV
                     AYCPHCDGPLFEGKDVAVIGGGNSGIEAAIDLAGIVNHVTVLEFAPELKADEVLQKRL
                     YSLPNVTVVKNAQTKEITGDQSVNGITYVDRETGEEKHVELQGVFVQIGLVPNTEWLE
                     GTVERNRMGEIIVDKHGATSVPGLFAAGDCTDSAYNQIIISMGSGATAALGAFDYLIR
                     N"
     gene            complement(2712. .4151)
                     /gene="bglA"
     CDS             complement(2712. .4151)
                     /gene="bglA"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="phospho-bete-glucosidase"
                     /protein_id="BAA11270.1"
                     /db_xref="GI:1064784"
                     /translation="MGNMPKDFLWGGALAAHQFEGGWNQGGKGPSVVDVMTAGAHGVP
                     RKITDTIEENEFYPNHEAIDFYHRYKEDIALFAEMGLKCLRTSIGWSRIFPKGDEAEP
                     NEAGLQFYDDVFDELLKHGIEPVITLSHFEMPLHLAREYGGFRNRKVVDFFVNFAEAC
                     FTRYKDKVKYWMTFNEINNQMDVNNPLFLWTNSGVVVGENENAKEVMYQTAHHELVAS
                     ALAVAKGKDINPEFQIGAMVSHVPIYPFSSNPEDVMLAEEEMRQRYFFPDVQVRGYYP
                     SYALKEFEREGYNITFEDGDDEILRNGTVDYLGFSYYMSTTVKSDVKNDNTGDIVNGG
                     LPNGVENPYITSSDWGWAIDPTGLRYTLNRFYDRYQIPLFIVENGFGAVDTLEEDGKV
                     HDPERIQYLKSHIEALKKAVTYDGVDLIGYTPWGIIDIVSFTTGEMKKRYGMIYVDRD
                     NEGNGSMKRYKKDSFEWYKNVIQTNGEEL"
     gene            4733. .5449
                     /gene="yydK"
     CDS             4733. .5449
                     /gene="yydK"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11271.1"
                     /db_xref="GI:1064785"
                     /translation="MRMLKYQQIATEIETYIEEHQLQQGDKLPVLETLMAQFEVSKST
                     ITKSLELLEQKGAIFQVRGSGIFVRKHKRKGYISLLSNQGFKKDLEDFNVTSKVIELD
                     VRKPTPEAAENLNIGMDEDIYYVKRVRYINGQTLCYEESYYTKSIVTYLNNEIVSHSI
                     FHYIREGLGLKIGFSDLFLHVGQLNEEEAEYLGLEAGLPKLYIESIFHLTNGQPFDYS
                     KISYNYEQSQFVVQANSFLL"
     gene            complement(5766. .6488)
                     /gene="yydJ"
     CDS             complement(5766. .6488)
                     /gene="yydJ"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11272.1"
                     /db_xref="GI:1064786"
                     /translation="MKLEFKKSISNKVIIILGAMFVFLFLLGYFLLVGIDKVSNVTPE
                     MFFFSSYTVATQFGLMLFSFVIAFFINREYSNKNILFYKLIGENIYTFFYKKIAVLFL
                     ECFAFITLGLLIISLMYHDFSHFALLLFLFSAVILQYILIIGTISVLCPNILISIGVS
                     IVYWMTSVILVAISNKTFGFIAPFEAGNTMYPRIERVLQSDNMTLGSNDVLFIILYLV
                     SIIIINAIVLRFSKTRWIKMGL"
     gene            complement(6509. .7138)
                     /gene="yydI"
     CDS             complement(6509. .7138)
                     /gene="yydI"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11273.1"
                     /db_xref="GI:1064787"
                     /translation="MNIANYTLKVKGKTLLQDTDLHFSSGKINHVVGKNGVGKSQLAK
                     DFLLNNSKRIGRDIRQNVSLISSSSNIPNDVSKDFLLHFLSKKFDAKMIDKIAYLLNL
                     DNIDGKVLIKNLSDGQKQKLKLLSFLLEDKNIIVLDEITNSLDKKTVIEIHGFLNKYI
                     QENPEKIIINITHDLSDLKAIEGDYYIFNHQEIQQYHSVDKLIEVYINE"
     gene            complement(7288. .8046)
                     /gene="yydH"
     CDS             complement(7288. .8046)
                     /gene="yydH"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11274.1"
                     /db_xref="GI:1064788"
                     /translation="MKILKYEDGKYEVLVQNNVFIKDKKSGEYYKNSLNSLSDKQLLR
                     FKMYKEKVSPKFFYLFLSFTALMFILNYIHLIKLQNGLSSVFYGWKMWIIIVIYFIMN
                     IVLHELGHIYSLKFFGKNFDKVGFKLNFYVFPAFYVQLNETYMLSRNEKIIVHLFGLF
                     INYLLINTLELINQFTFSSEALTMAFMLFSSTLLWNLIPILNSDGYKILLAFLSLDEY
                     SRFKTNHWLVLTIQIIGIGLAVNSVVHWILYIVN"
     gene            complement(8027. .8986)
```

```
                /gene="yydG"
     CDS        complement(8027. .8986)
                /gene="yydG"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA11275.1"
                /db_xref="GI:1064789"
                /translation="MYNKTVSINLDSRCNASCDHCCFSSSPTSTTRMEKEYIRELVTE
                FAKNKTIQVISFTGGEVFLDYKFLKELMEIIKPYEKQITLISNGFWGLSKKKVQEYFH
                DMNSLNVIALTISYDEYHAPFVKSSSIKNILEHSRKYPDIDISLNMAVTKDKMSNHIL
                EELGDSILGVKITKFPMISVGAAKTRIKQENIHKFYSLEDEDSLHCPGYDIVYHHDGE
                IYPCCSPAIFETKITLREEYNQSFERTVEKLNSNLLLFILRKEGFKWFLNILKENNKI
                EEFDIPYEFSSICGVCGSLFNSAEKINYFYPYMEKYYNENFKV"
     gene       complement(9044. .9193)

  Query Match             3.4%;  Score 46;  DB 1;  Length 36200;
  Best Local Similarity   60.3%;  Pred. No. 1.6;
  Matches   76;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      51 gcaagccgtttacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgc 110
           ||   | |  | || || |       || |  |  ||| | ||  |  |    || ||
Db    4363 GCGTTCAGCGTCCGGCGTCGAATTGCTTGCGCACATCGGGCTGGATACGGTGACATTTGA 4304

Qy     111 cggtatgcgccccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgct 170
           ||| | ||     ||  |||| || || |||||||||| |||||||||||||| |||||||
Db    4303 CGGCACGCCTTTTTCCTTGAAGGTAAAAGAAGGCGATACCGTCAAAAAAGGCGAAGTGCT 4244

Qy     171 gtttga 176
              ||||
Db    4243 TGTTGA 4238


RESULT  22
BSUB0021/c
LOCUS       BSUB0021              215534 bp    DNA     linear   BCT 26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 21 of 21): from 3999281
            to 4214814.
ACCESSION   Z99124 AL009126
VERSION     Z99124.1  GI:2636442
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 215534)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 215534)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1. .215534
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            complement(11. .820)
                     /gene="yxjF"
     terminator      complement(11. .37)
                     /gene="yxjF"
     CDS             complement(47. .820)
                     /gene="yxjF"
                     /function="unknown"
                     /note="similar to gluconate 5-dehydrogenase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15933.1"
                     /db_xref="GI:2636443"
                     /db_xref="SWISS-PROT:P42317"
                     /translation="MRKQVALVTGAAGGIRFEIAREFAREGASVIVSDLRPEACEKAA
                     SKLAEEGFDAAAIPYDVTKEAQVADTVNVIQKQYGRLDILVNNAGIQHVAPIEEFPTD
                     TFEQLIKVMLTAPFIAMKHVFPIMKKQQFGRIINIASVNGLVGFAGKSAYNSAKHGVI
                     GLTKVGALEGAPHGITVNALCPGYVDTQLVRNQLSDLSKTRNVPYDSVLEQVIFPLVP
                     QKRLLSVKEIADYAVFLASEKAKGVTGQAVVLDGGYTAQ"
     gene            complement(837. .1487)
                     /gene="yxjE"
     CDS             complement(837. .1487)
                     /gene="yxjE"
                     /function="unknown"
                     /note="similar to 3-oxoadipate CoA-transferase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15934.1"
                     /db_xref="GI:2636444"
                     /db_xref="SWISS-PROT:P42316"
                     /translation="MKEARKRMVKRAVQEIKDGMNVNLGIGMPTLVANEIPDGVHVML
                     QSENGLLGIGPYPLEGTEDADLINAGKETITEVTGASYFDSAESFAMIRGGHIDLAIL
                     GGMEVSEQGDLANWMIPGKMVKGMGGAMDLVNGAKRIVVIMEHVNKHGESKVKKTCSL
                     PLTGQKVVHRLITDLAVFDFVNGRMTLTELQDGVTIEEVYEKTEADFAVSQSVLNS"
     gene            complement(1484. .2200)
                     /gene="yxjD"
     CDS             complement(1484. .2200)
                     /gene="yxjD"
                     /function="unknown"
                     /note="similar to 3-oxoadipate CoA-transferase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15935.1"
                     /db_xref="GI:2636445"
                     /db_xref="SWISS-PROT:P42315"
                     /translation="MGKVLSSSKEAAKLIHDGDTLIAGGFGLCGIPEQLILSIRDQGV
                     KDLTVVSNNCGVDDWGLGLLLANKQIKKMIASYVGENKIFERQFLSGELEVELVPQGT
                     LAERIRAGGAGIPGFYTATGVGTSIAEGKEHKTFGGRTYVLERGITGDVAIVKAWKAD
                     TMGNLIFRKTARNFNPIAAMAGKITIAEAEEIVEAGELDPDHIHTPGIYVQHVVLGAS
                     QEKRIEKRTVQQASGKGEAK"
     gene            complement(2279. .3643)
                     /gene="yxjC"
     CDS             complement(2279. .3643)
                     /gene="yxjC"
                     /function="unknown"
```

```
                /note="similar to hypothetical proteins from B. subtilis"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15936.1"
                /db_xref="GI:2636446"
                /db_xref="SWISS-PROT:P42314"
                /translation="MELIIILLALGLLMFTAYRGFSVILFAPICALFAVLLTDPSHVL
                PFFSSIFMEKMAGFIKLYFPVFLLGAIFGKVVEMAGLAASIAKTIVRLVGAKRAILAI
                VLMGAVLTYSGVSLFVVVFAVYPFAKNMFQEANIPKRLIPGTIALGAFTFTMDALPGT
                PQIQNVIPTSFFKTDIYAAPWLGLMGAVIVLAAGMLYLESRRKKAQASGEGYGGFDSQ
                NAPAPESIESAAEPDKSPIRHALAFVPLILVGAVNKYFTIYLPKWYPNGFNFPSIGLK
                EFGRLDISSAAAIWSVEIALVIGIITTILFDWRSVFAQLKEGLNEGIGGALLASMNTG
                AEYGFGGIIAALPGFHKLSSGISHTFTDPLVNGAVTTTALAGITGSASGGMGIALSAM
                SEQYLQAIQAYNIPPEVMHRVISMASGGMDTLPHNGAVITLSGRDGFDPPAILSRYFC
                DHAH"
gene            complement(3762. .4645)
                /gene="yxjB"
terminator      complement(3762. .3788)
                /gene="yxjB"
CDS             complement(3797. .4645)
                /gene="yxjB"
                /function="unknown"
                /note="similar to hypothetical proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15937.1"
                /db_xref="GI:2636447"
                /db_xref="SWISS-PROT:P42313"
                /translation="MKRTVDFSMFRCPLCDSSMDAASGKSLICTERGHTFDLSRHGYV
                NFLTKPVKTSYGAELFEARSRLIGECGFFDPLHDAIAELISHPKSGHEAFTILDSGCG
                EGSHLNALCGFDYAGKAAIGTGIDLSKDGILKASKAFKDLMWAVADVARAPFHDRQFD
                VVLSIFSPSNYAEFHRLLKNDGMLIKVVPRSDYLIELRQFLYTDSPRRTYSNTAAVER
                FTANAAHSRPVRLRYVKTLDQQAIHWLLKMTPLAWSAPKDRVSLLKEMKSADITVDVD
                ILIGMK"
gene            5261. .6489
                /gene="yxjA"
CDS             5261. .6454
                /gene="yxjA"
                /function="unknown"
                /note="similar to pyrimidine nucleoside transport"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15938.1"
                /db_xref="GI:2636448"
                /db_xref="SWISS-PROT:P42312"
                /translation="MYFLLNLVGLIVIMAVVFLCSPQKKKIKWRPIITLIVLELLITW
                FMLGTKVGSWAIGKIGDFFTWLIACASDGIAFAFPSVMANETVDFFFSALLPIIFIVT
                FFDILTYFGILPWLIDKIGWVISKASRLPKLESFFSIQMMFLGNTEALAVIRQQLTVL
                SNNRLLTFGLMSMSSISGSIIGSYLSMVPATYVFTAIPLNCLNALIIANLLNPVHVPE
                DEDIIYTPPKEEKKDFFSTISNSMLVGMNMVIVILAMVIGYVALTSAVNGILGVFVHG
                LTNQTIFAYLFSPFAFLLGLPVHDAMYVAQLMGMKLATNEFVAMLDLKNNLKSLPPHT
                VAVATTFLTSFANFSTVGMIYGTYNSILDGEKSTVIGRNVWKLLVSGIAVSLLSAAIV
                GLFVW"
terminator      6462. .6489
                /gene="yxjA"
gene            complement(6925. .7167)
                /gene="yxiT"
CDS             complement(6925. .7167)
                /gene="yxiT"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15939.1"
                /db_xref="GI:2636449"
                /db_xref="SWISS-PROT:P42311"
                /translation="MKWNNMLKAAGIAVLLFSVFAYAAPSLKAVQAKTPTVSTHTYKK
                IKELTYPQVHHVGNAAFEKKINQELKAYMDNRIRNT"
terminator      complement(7290. .7306)
                /gene="yxiS"
gene            complement(7290. .7602)
                /gene="yxiS"
CDS             complement(7312. .7602)
                /gene="yxiS"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15940.1"
                /db_xref="GI:2636450"
                /db_xref="SWISS-PROT:P42310"
                /translation="MNRKQLTEAYERDENMMILVFAQWCVNHDLDPMELYAKAYPQQK
                LNESLKKTMDDLVVPKHEAEHIPDQTVIAVLEMFGNTDLAQALHEAISGRKQ"
gene            complement(7652. .9712)
                /gene="katB"
CDS             complement(7652. .9712)
                /gene="katB"
                /EC_number="1.11.1.6"
                /note="alternate gene name: katE"
                /codon_start=1

  Query Match             3.4%;  Score 46;  DB 1;  Length 215534;
  Best Local Similarity   60.3%;  Pred. No. 1.7;
  Matches   76;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      51 gcaagccgtttacgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgc 110
           ||   | |  | || || |       || |  |  ||| | ||  |   |    || ||
Db 122745 GCGTTCAGCGTCCGGCGTCGAATTGCTTGCGCACATCGGGCTGGATACGGTGACATTTGA 122686

Qy     111 cggtatgcgcccctcgatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgct 170
           ||| | ||      ||  |||| || || |||||||| |||||||||||||| |||||||
Db 122685 CGGCACGCCTTTTTCCTTGAAGGTAAAAGAAGGCGATACCGTCAAAAAAGGCGAAGTGCT 122626

Qy     171 gtttga 176
             ||||
Db 122625 TGTTGA 122620


RESULT  23
OSJN00193/c
LOCUS       OSJN00193             109730 bp    DNA     linear   HTG 27-DEC-2001
DEFINITION  Oryza sativa chromosome 4 clone OSJNBb0004G23, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL662992
VERSION     AL662992.1  GI:17998504
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,
            Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J., Zhang,L.,
            Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y., Li,C.,
            Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K., Zhou,B.,
            Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P., Fu,G.,
            Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F., Tu,Y.F.,
            Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H., Chen,X.Y.,
            Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W., Wang,L.J.,
            Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L., Zhu,F.H. and
            Hong,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-DEC-2001) Han Bin, National Center for Gene Research,
            Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
            clone: OSJNBb0004G23.
COMMENT     Web site: http://www.ncgr.ac.cn
            ------------- Summary Statistics
            Assembly program: phrap
            -------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES              Location/Qualifiers
```

```
     source          1. .109730
                     /organism="Oryza sativa"
                     /variety="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:4530"
                     /chromosome="4"
                     /clone="OSJNBb0004G23"
                     /clone_lib="CUGI-OSJNBb"
BASE COUNT     30967 a  24013 c  24332 g  30355 t     63 others
ORIGIN


  Query Match             3.1%;  Score 41.4;  DB 2;  Length 109730;
  Best Local Similarity   47.2%;  Pred. No. 22;
  Matches  126;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    1026 gccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaactctt 1085
           | |||  || | ||||| |          | | | | |       |||  | ||  |||| 
Db   87723 GGCGGCTAATTCCTCCAACGACACCGTCAACGTGGCCTCCGAGGTGAGCAGCATCCTCTC 87664

Qy    1086 caagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtacttacga 1145
           ||||  |||||     | ||  ||||| ||||   |||| || ||||   || ||   | 
Db   87663 CAAGCTCAACGATCACCTCGCCGGCGCCGACGAGGCCAAGGAGCCGGCGGGGACAAGCAT 87604

Qy    1146 gcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtcggcga 1205
             |  | || ||  ||| | |     | ||    |   |       ||| |  ||   
Db   87603 CATCACGCTGGCCGGGGAGAACAACGGCGCCACCATGGAGGTGGCCGGCGACGTGGAGGA 87544

Qy    1206 taccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgctttgtg 1265
             || |   | | ||||    ||   |  |||    ||||| || ||       ||  
Db   87543 CCTCGTCGTGGTGGAGGCCGGCGGCGACGAGGACGACGACGAGGAGGAGGAGAGCGTGGT 87484

Qy    1266 cagcttcgtctgcccgggcaaatacga 1292
            ||   |  |  |    |||| ||| | 
Db   87483 GAGTGCCTACACCAACAGCAACTACCA 87457


RESULT  24
OSJN00146
LOCUS       OSJN00146             142805 bp    DNA     linear   HTG 27-DEC-2001
DEFINITION  Oryza sativa chromosome 4 clone OSJNBa0094O15, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL662935
VERSION     AL662935.1  GI:17998447
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Fu,G., Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F.,
            Tu,Y.F., Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H.,
            Chen,X.Y., Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W.,
            Wang,L.J., Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L.,
            Zhu,F.H., Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q.,
            Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J.,
            Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y.,
            Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K.,
            Zhou,B., Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P.
            and Hong,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-DEC-2000) Han Bin, National Center for Gene Research,
            Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
            clone: OSJNBa0094O15.
COMMENT     Web site: http://www.ncgr.ac.cn
            ---------- Summary Statistics
            Assembly program: phrap
            ----------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .142805
                     /organism="Oryza sativa"
                     /variety="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:4530"
                     /chromosome="4"
                     /clone="OSJNBa0094O15"
                     /clone_lib="CUGI-OSJNBa"
BASE COUNT     42656 a  28666 c  28764 g  42718 t      1 others
ORIGIN


  Query Match             3.1%;  Score 41.4;  DB 2;  Length 142805;
  Best Local Similarity   47.2%;  Pred. No. 23;
  Matches  126;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    1026 gccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaactctt 1085
           | |||  || | ||||| |          | | | | |       |||  | ||  |||| 
Db  132672 GGCGGCTAATTCCTCCAACGACACCGTCAACGTGGCCTCCGAGGTGAGCAGCATCCTCTC 132731

Qy    1086 caagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtacttacga 1145
           ||||  |||||     | ||  ||||| ||||   |||| || ||||   || ||   | 
Db  132732 CAAGCTCAACGATCACCTCGCCGGCGCCGACGAGGCCAAGGAGCCGGCGGGGACAAGCAT 132791

Qy    1146 gcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtcggcga 1205
             |  | || ||  ||| | |     | ||    |   |       ||| |  ||   
Db  132792 CATCACGCTGGCCGGGGAGAACAACGGCGCCACCATGGAGGTGGCCGGCGACGTGGAGGA 132851

Qy    1206 taccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgctttgtg 1265
             || |   | | ||||    ||   |  |||    ||||| || ||       ||  
Db  132852 CCTCGTCGTGGTGGAGGCCGGCGGCGACGAGGACGACGACGAGGAGGAGGAGAGCGTGGT 132911

Qy    1266 cagcttcgtctgcccgggcaaatacga 1292
            ||   |  |  |    |||| ||| | 
Db  132912 GAGTGCCTACACCAACAGCAACTACCA 132938


RESULT  25
AP003456
LOCUS       AP003456              151329 bp    DNA     linear   HTG 29-MAR-2001
DEFINITION  Oryza sativa chromosome 6 clone P0542E10, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP003456
VERSION     AP003456.1  GI:13486925
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0542E10.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
            clone:P0542E10
  JOURNAL   Published Only in DataBase (2001) In press
REFERENCE   2  (bases 1 to 151329)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
```

```
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .151329
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="6"
                     /clone="P0542E10"
BASE COUNT    40598 a  33769 c  34863 g  41449 t    650 others
ORIGIN


  Query Match             3.0%;  Score 40.2;  DB 2;  Length 151329;
  Best Local Similarity   46.3%;  Pred. No. 45;
  Matches  132;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy    1059 cggccatttcctgaaaaacaaactcttcaagttcaacacagccgtcaacggcggcgaccg 1118
           || ||  ||| | ||   ||| ||||||||    || | | | ||||  || |  |  |
Db   25301 CGCCCTCTTCTTCAACCTCAACCTCTTCAGCGGCATCTCCGACGTCGGCGCCATCCTCGA 25360

Qy    1119 cgccatggtgccgattggtacttacgagcgcgtgatgcccttggatatcctgcccaccct 1178
           | ||| ||| |   |      | | ||    |   || |||      | |     ||
Db   25361 CCCCAAGGTCCGCGTCATCCTCTCCAAGGCGCTCTCGCTCTTCCTCCCCGTCATGTCCTA 25420

Qy    1179 gcttttgcgcgatttaatcgtcggcgataccgacagcgcgcaggcattgggttgcttgga 1238
            || ||   |||    |    || || | ||| |  | |    ||    ||| ||  ||
Db   25421 CCTCTTCTCCGAGGCCAAGAACGCCGGTGCCGCCGCCTCCGCCGCCGGCGGTAGCACGGG 25480

Qy    1239 attggacgaagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacgg 1298
            | |  |||    || ||| |  || |        |   ||  |             ||
Db   25481 CTCGCTCGAGCTCGAGCTCTCGCTGCGGGCGCGGCTGATCCTCGTGTGGATGCTCCTCGT 25540

Qy    1299 cccgctgttgcgcaaagtgctggaaaccattgagaaggaaggctg 1343
           |  |||| | |||||   | ||||    |||  || ||| |  ||
Db   25541 CGAGCTGCTCCGCAAGAAGGTGGAGGAGATTCGGATGGAGGCATG 25585


RESULT  26
AL646076
LOCUS       AL646076               216050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
            segment 1/11.
ACCESSION   AL646076 AL646053
VERSION     AL646076.1  GI:17430467
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 216050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .216050
                     /organism="Ralstonia solanacearum"
                     /plasmid="megaplasmid"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     rep_origin      1. .17
                     /note="ori or RS05845
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   52. .68
                     /note="RS05846
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(77. .93)
                     /note="RS05847
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   159. .175
                     /note="RS05848
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   202. .218
                     /note="RS05849
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(227. .243)
                     /note="RS05850
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(263. .279)
                     /note="RS05851
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(307. .323)
                     /note="RS05852
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(374. .390)
                     /note="RS05853
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(516. .532)
                     /note="RS05854
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(670. .686)
                     /note="RS05855
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(733. .749)
                     /note="RS05856
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(853. .869)
                     /note="RS05857
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
```

```
                /evidence=not_experimental
repeat_region   complement(919. .935)
                /note="RS05858
                HYPOTHETICAL PARA BOX
                predicted by Homology"
                /evidence=not_experimental
gene            1219. .2244
                /gene="repA"
                /note="RSp0001; RS02238"
CDS             1219. .2244
                /gene="repA"
                /function="elements of external origin; plasmid-related
                functions non-bacterial functions"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE REPA REPLICASE PROTEIN"
                /protein_id="CAD17152.1"
                /db_xref="GI:17430468"
                /translation="MPGRIIVPSGGQVIATPEKFQRLFDEAQAMEREDAWQSGEVGFL
                SRASVQVTLPYRAPKGSPPVWTRSSGNISLMIQPGYFTQQRSERASNGRQRIVSETVS
                FGYPYGSYPRLMLAWIGKEIMAKKKRGEFQGTVEDRRISLGNSLSEFMYNLGIPMATG
                GKRGTMTLVRQQMIRLFSATIAIVQNKSAPNNQNQEPLSIDRIGYLLADQLSTWWDPM
                QPGQGSMFESFVVLSEPFFNELVNRPVPVDMRALKALKQSPFALDVYSWLTYRFFTIQ
                KRTEIPWEALQMQFGTETESERKFRALFRKALKDVLVVYPDAKVDADSSKALILQPSR
                TSVRKLA"
gene            complement(3120. .3323)
                /gene="cspD3"
                /note="RSp0002; RS05790"
CDS             complement(3120. .3323)
                /gene="cspD3"
                /function="miscellaneous; not classified regulator"
                /note="Product confidence : probable
                Gene name confidence : putative
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE COLD SHOCK-LIKE TRANSCRIPTION REGULATOR
                PROTEIN"
                /protein_id="CAD17153.1"
                /db_xref="GI:17430469"
                /translation="MATGTVKWFNETKGFGFITPDGGGADLFAHFSEIQGSGFKTLKD
                GQKVSFEVKQGPKGLQASGIKPE"
gene            3728. .4396
                /gene="parA"
                /note="RSp0003; RS01975"
CDS             3728. .4396
                /gene="parA"
                /function="elements of external origin; plasmid-related
                functions non-bacterial functions"
                /note="Product confidence : probable
                Gene name confidence : putative
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE PARTITION PROTEIN"
                /protein_id="CAD17154.1"
                /db_xref="GI:17430470"
                /translation="MSAKIVTVFNQKGGCGKTTVSMHIAGTLGVRGSKTLLVDMDEQG
                TATRWAAQAPEEKPFPASVIGLAPSGGAMHREVRKFIADYDYIVVDCPPAVHSPASSS
                ALLISDIAIIPVVPSPPDLWAAVAAKALAQQAQVTNETLRVRVLANMVQRRVSLAKQA
                IEILGDDGDIPLMESMIGSRSAFRECQAIGATVHGVPGAREAINEVDIMVNEVLQLLR
                EGQA"
gene            4393. .5397
                /gene="parB"
                /note="RSp0004; RS01976"
CDS             4393. .5397
                /gene="parB"
                /function="macromolecule metabolism; macromolecule
                synthesis, modification; dna - replication, repair,
                restr./modif."
                /note="Product confidence : putative
                Gene name confidence : putative
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PUTATIVE PARTITIONING PROTEIN"
                /protein_id="CAD17155.1"
                /db_xref="GI:17430471"
                /translation="MSAKLKSLKDGMLAGIAAEKTKAAPVDRFALAEQAITSHPRGLL
                GQPARPAFSAESQAETSSERRVVKIPLERLQENPLNARRIYDQQIVQERAASIATHGQ
                QTPGLAAPDPAKPGWYVLIDGHYRKRALAAAGKAEMECFIEDGLSDIDFYRLSFVLNE
```

```
  Query Match             3.0%;  Score 39.8;  DB 1;  Length 216050;
  Best Local Similarity   48.5%;  Pred. No. 57;
  Matches  110;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    1085 tcaagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtacttacg 1144
           || |   |||  | | || || || || | ||| | ||   |         | |    ||
Db   20800 TCGACCACAAGGCGGGCGACATCGCCGACAACCTCTCCTATGGCGACCAGCGCAAGCTCG 20859

Qy    1145 agcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtcggcg 1204
           ||| || |  | | ||||  | |  ||| |  |||||  ||| |||      ||||| ||
Db   20860 AGCTCGCGCGGGCGTTGGCCACCGAGCCGAAGCTGCTGCTGCTCGACGAGCCCGTCGCCG 20919

Qy    1205 ataccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgctttgt 1264
             |  ||| ||   ||    | | |    |||    |  |||  |   ||  ||
Db   20920 GCATGAACACCGGTGAGAAGGTCGATCTGATGGCCGAGATCGAGAACATCCGGGCGCGCG 20979

Qy    1265 gcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgc 1311
           ||  |  | ||| |  |  | |  ||||     ||    || || ||
Db   20980 GCTACACCATCTTCATGATCGAGCACGACATGCGCTTCGTGATGGGC 21026
```

```
RESULT  27
AB070947/c
LOCUS       AB070947               24268 bp    DNA     linear   BCT 11-OCT-2001
DEFINITION  Streptomyces avermitilis polyketide-9 biosynthetic gene cluster.
ACCESSION   AB070947
VERSION     AB070947.1  GI:15824107
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 24268)
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: Deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
   PUBMED   11572948
REFERENCE   2  (bases 1 to 24268)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
```

```
FEATURES             Location/Qualifiers
     source          1. .24268
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
                     /note="Actinobacteridae"
     CDS             1. .1875
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="BAB69274.1"
                     /db_xref="GI:15824108"
                     /translation="MSAYHFSPWIAVAVVVLMFTVVASYRQNVHAYPSGGGDYEVANT
                     NLGPRAGLTVASALLVDYVLTVAVSIASGIENLGSAIPFVVEHKVSCAVAVIVLLTLM
                     NLRGVKESGKLFAIPTYVFVAGVFIMIAWGAFRGFVLDDTMHAPTADYHIKAEHQGLA
                     GFALVFLLLRAFSSGCAALTGVEAISNGVPAFRKPKSKNAASTLAAMGLLAVTMFCGI
                     IALAMATNVRMAENPANDLIHNGVAVGPDYVQNPVISQVAEAVFGKGSFLFIVLAAAT
                     ALVLFLAANTAYNGFPLLGSILAQDRYLPRQLHTRGDRLAFSNGIVLLAGAATLLVVI
                     YGADSTRLIQLYIVGVFVSFTLSQTGMVRHWNRHLATEKDQAKRRHMVRSRAINAFGA
                     FFTGLVLVVVLVTKFTHGAWVALLGMCIFYATMSAIRKHYDRVAEEIAAPEGPSDDSV
                     RPSRVHSVVLISKIHRPTLRALAYAKLMRSDTLEALSVNVDPAETKALKEEWDRRGID
                     VPLKVLDSPYREITRPVIEYVKGLRKESPRDAVSVIIPEYVVGHWYEHLLHNQSALRL
                     KGRLLFTPGVMVTSVPYQLESSEAAKKRARRRQEWSAPGAVRRGPAEERAKESSNTKG
                     "
     CDS             1995. .3296
                     /codon_start=1
                     /transl_table=11
                     /product="putative methyltransferase"
                     /protein_id="BAB69275.1"
                     /db_xref="GI:15824109"
                     /translation="MGEEYEVEIGPVAHGGHCIARTSEGQVLFVRHALPGERVLARVT
                     EGEEGARYLRADAVEILDASKDRVEAPCPYAGPGRCGGCDWQHAKPGAQRRLKGEVIA
                     EQLQRLAGLTPEEAGWDGTVMPAEGDKLPAGEVPAWRTRVQYAVDADGNAGLRRHRSH
                     EVEPIEHCMIAAPGVSELGIEERDWSGMESVDAIAATGSQDRMVILEPRPGARLPLVE
                     LDKPVSVMRVEEKDGGIHRVHGRAFVRERADGRTYRVGSGGFWQVHPMAADTLVKAVM
                     QGLLPRKGDMALDLYCGVGLFAGALADRLGDKGAVLGIESGKRAVEDARHNLAAFERV
                     RIEQGKVEAVLPRTGITEVDLIVLDPPRAGAGKKTVEQLVSLGARKIAYVACDPAALA
                     RDLGYFRDGGYKVRTLRAFDLFPMTHHVECVAILEPAAKGL"
     CDS             complement(4334. .5074)
                     /codon_start=1
                     /transl_table=11
                     /product="3-oxoacyl-(acyl carrier protein) reductase"
                     /protein_id="BAB69276.1"
                     /db_xref="GI:15824110"
                     /translation="MATPGTDGRVALVTGASGALGAAVARVLDHRGYRLALHYTNGED
                     QARKVAAKLSGPSTLVRADVADWAATRDMVDRVRAELGGISVLVNTGAIRRDGLMATQ
                     SVDDWTRTVAVNLIGTFHTCRAALPDMLRGRWGRIINVVSPAGLLGSRGQTAYSASKA
                     GVLGMTRSLALECGRFNVTVNALSPGLMATALTEEVPAEVRATLVARTAFGRMGTPDE
                     VARGIELLLDADYMTGQVLSIDGGMSIS"
     gene            complement(5513. .6217)
                     /gene="clpX"
     CDS             complement(5513. .6217)
                     /gene="clpX"
                     /codon_start=1
                     /transl_table=11
                     /product="ClpX homolog"
                     /protein_id="BAB69277.1"
                     /db_xref="GI:15824111"
                     /translation="MTQITPPVKLDELITAITENHTTALDQLSGAILLADHIGEVADH
                     LIGHFVDQARHSGASWTDIGRCMGVTRQAAQKRFVGKPDAKPSNAFEKFSQPARSAII
                     AGMNEAKTAGNAEIVPAHVLLGLLGTEGSAVGALRAQGIDLDQARDTAVATLPPRSDA
                     DPVLIPYDAHAKKALELASRVAFRLGQDSVSTGHMLLGLLEEGDGAGVLASLGLDTAA
                     AERFVAEAADESSVAS"
     CDS             complement(6353. .7513)
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P450"
                     /protein_id="BAB69278.1"
                     /db_xref="GI:15824112"
                     /translation="MLPAEYETLRERGIGQVELPTGKLVWMVVRPEYARIVLSDPRFS
                     SDKTDDRFPKLTPNSLMKLRYCAPFMINLDGPEHLKKKKSIMDEFSPEGLARLLPRLR
                     VAVEEKIDDMLRQPTKPVDLVKELAFPIAWQLQEMFLGIPAAELETMRDNVWKLLLGT
                     TTEAEEREAADRLNGHAEEVLKEKAKHLGDDMMSRLIVQEREKHGEVNWYELAPLMLS
                     NAQGIHNSVSTMISLGVLTLLNHPEQRPTLLAHPDRMTTAVDEMLRYFSVNDGTPMRL
                     ATEDMLIGDTLVKAGDGVAVPTLPVNRDPSVCPYPHQLDIMREEPARHLAFGHGPHKC
                     PADRLVPSLLEIVYTTLFERVPTLALAVPEAELTYKYHSIQAFGPAEMPVTW"
     CDS             complement(7555. .8232)
                     /codon_start=1
                     /transl_table=11
                     /product="cyclase"
                     /protein_id="BAB69279.1"
                     /db_xref="GI:15824113"
                     /translation="MPSIITVGAFDSGNVADVTKILQEFDGEVASELPQLRRRQVFSY
                     SDLYVHMQDWDVADPDEAYRIISTDPRCTRLNERLSTYFGDHAATEDWDGPVDRPTAE
                     RFYLWPAERLDDLEQTYSAVIVNTQRQEHIPEAARLFAESDATDIPQKMGTLRRQIYL
                     WRGIYLHIQDFAVANSMEAISEVWTEGDPRFLQLVDDLTEIIPPYDPQGNSLATRFYH
                     WAAEVPA"
     CDS             8414. .9382
                     /codon_start=1
                     /transl_table=11
                     /product="aromatase"
                     /protein_id="BAB69280.1"
                     /db_xref="GI:15824114"
                     /translation="MTTLVRCLVVSSTQRRVDHAITIEGDVETVFDLIADIARWPQFH
                     RSAVHAETVHANEKGELVRHWALAGDAAVRAWQTVRRTDRPGRRITFAHMEPEQPATD
                     LRGEWVFTELTPNRTRVELSHQVVAPTGADLGPLLDRIAQESEDCLATLRDTAERRAD
                     LDELVVSFTDPLFIAGSVEDAYDYLYRADLWPERIPHVVGLAMTEDVPNVQFFDMDTK
                     GPDGSTHTVRSVRICLPNHLIVYKQISLPTMLDAHTGHWRFTETPEGVIAEARHTSTI
                     KRSALSQLGGATTVAEARRYLRRLLSGNSMSNLRLAKAYAEERAGF"
     CDS             9449. .10528
                     /codon_start=1
                     /transl_table=11
                     /product="O-methyltransferase"
                     /protein_id="BAB69281.1"
                     /db_xref="GI:15824115"
                     /translation="MSCRTGTDTVPAGSHEQRTVESGEVMAKETTPRGGGVWAAADLL
                     TPMAVRVAATLRLADHIAAGARTTEALAEAVGADRDALGRLLDHLVTAGVLSGTGPGA
                     YDLTAMGRHLCEGAPEDMRAILDIEGALGHAELSLVHLLHTVRTGEAAFPQQYGVTFW
                     DDLSSDDGRAESFDTLMGARLTAHSPAVAGAYPWGTLRHVVDVGGGDGTMLIAILQSH
                     PDLRGTVVDLPGPVRRAEKAIAAAGLDHRADIAAGSFFDALPAGADGYLLSSILHNWD
                     DASAARILRRCADAAQTTGRVLVVDYFGDRTVQTEGDLRMLGYFGGRQHTLEQLAELA
                     GTVGLHTTSVTPAGRYSVVELRAVG"
     CDS             complement(10495. .11280)
                     /codon_start=1
                     /transl_table=11
                     /product="tetR-family regulatory protein"
                     /protein_id="BAB69282.1"
                     /db_xref="GI:15824116"
                     /translation="MRPEKPARGPAPSHSRRQIAAVAIDIADSEGIEAASVRAIAKRL
                     GTGAMTLYRYLPTKEDLYTVMIDEATGFEPQEPTGDLRADLTTLARRRRQILLRHLWL
                     APLLATRPIMGPNFMRGMERDLAVLAGCGLGMDDTIHVLNLIHSWVSGAVQAELTERA
                     TANQSGLDRHAWRLRMEPYLKSLLETGEFPHLSRMVQTSEIGEADERFANGLAIILDG
                     IEARFPTATGKDGSRRPVGDTGGGAPSARRAKGQPTARSSTTE"
     CDS             11434. .12375
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter I"
                     /protein_id="BAB69283.1"
                     /db_xref="GI:15824117"
                     /translation="MRKQYGDVVALDGLDLVVPSRTVYGLLGPKGAGKSTAVRILATL
                     TRPSSGRAVVAGFDVARNPADVRKVIGLAGQDASFDDGLTGRENLRILGRLSRLGSRG
                     ARRRADELLERFGLAHAGDRLVGTYSGGMRRRIHLIASLLRAPAVLFLDEPTTGLDAP
                     SRAEIWESVRRLVADGTTVLLTTPCLEEVERLATAVGVIDRGQLIAEGSPAQLKAQIG
                     RHIDVVVESADHVPSAMARLERLTGRCPEADGSRIVVAAVDQAPALPLIVSELDASGA
                     VIRNIGIREPGLDDVFLTMTGQTVGPAQPVETPEAVS"
     CDS             12372. .13157
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter II"
                     /protein_id="BAB69284.1"
                     /db_xref="GI:15824118"
                     /translation="MNLLADGWTLTQRRLFRLRHEPGTLLTMLMMPSGFVVLFGFVFG
                     GAIGVPGTSNYREFMMPGLFVMTTGTALSTAMVEVATDQARGVMERLRSLPTHRVSVP
                     LGQSGAEGILGAYGLAVLMVCGLVVGWRPHDGVPRALAGIGLLLLFRYTLAWVGTYLG
                     LVVRSARTAARLAPLTLSLTMVSNALVPTDRMPVGLRALCEWNPLSSAAAAARELFGN
                     PGAPGPDAAWPLTHPVTASLLWAGLLLLIFVPLSVRRFARAGL"
     CDS             13225. .14730
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="putative oxidase"
                     /protein_id="BAB69285.1"
                     /db_xref="GI:15824119"
                     /translation="MRRTREENTVLILGGTSMRVVIVGGGTVGLCTAVFLARQGVSVH
                     VLERRAQVNDHPRAFGITDRSAEILREAGIDIETDHLVRGLMTGTSLADLTTVTAKFE
                     PPEIGPAKLGSCPQDRVDRAALARLRELGGQVDFDSEVVAVTDGGVTTADGRTVTADY
                     VIAADGVRSRMRDLLGIGVTGPGPLSSSHMLNVLFRAELPTPPTSMTFLRNERASGIL
                     LKVDETGRWVFHIPNGDPETPVERIQQQIRDATGVSDIEPEIISLLPWSSTARVADRF
                     RQGNVFLVGDAAHVVPPVGGLGLNVGIADGHNLAWKLAAGGDALLDTYEAERRPVALF
                     TMRQTVLRAEHRDLHWDNDASRAADRAAVGMAGIMIPVIGYRYDSGAVVDPEPLSDME
                     NLVLDGAPGSRLPHRWLPDGRSTVDLVDGFTLITGTAADHWLDATRDLGITGHAVEGW
                     AEAVGIKEDGAVLVRPDGFVGWRTATTRNAVETSTQHLRDVLDRILCRQKA"
     CDS             15054. .16253
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P450"
                     /protein_id="BAB69286.1"
                     /db_xref="GI:15824120"
                     /translation="MDFGADPYPEYAWLRAEEPVRQVLEGRGLYGLLVTRYEDVRKLL
                     SDPRMSKDPRNAPLDWQEAGKGRPLEDRTGLGTHLLTTDAPEHTRLRRLVSTAFTARR
                     VEGLRAQVQHITDGLLDTIVPRGQAELIGDFAFPLAITVICELLGVPKADQDVFRQWT
                     KDFRRWTNTDSAQADRGDARPVGLRDLLEYLTRLVDKRRQDPADGLVDALIAARDDDD
                     RLNEAELLSMMSLLLVGGFETTVNLIGNGTLALLRHPDQLALLRERPELVDSALEEML
                     RYDGSFETATWRFPLEPIEVAGTRIEKGHPVLLSLASANRDGAKFPAPDDFDVTRADP
                     AHVAFGRGAHFCLGAPLARLEGRIAFHGLLRRLPGLALSVPPEQLRWQRSLTVRGLEA
                     LPVTFDA"

  Query Match              2.9%;  Score 39.2;  DB 1;  Length 24268;
  Best Local Similarity   51.7%;  Pred. No. 72;
  Matches   89;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     326 cgctggcaaacttaagcggcgaagaagtgcgccgcaacctgatccaatccggtttgtgga 385
           || || | |   | | |       |  |||| |  |   |||| |||  | |    | ||
Db    6837 CGGTGTCCACGATGATCTCGCTGGGCGTGCTCACCCTGCTGAACCACCCGGAGCAGCGGC 6778

Qy     386 ctgcgctgcgcacccgtccgttcagcaaaattcctgccgtcgatgccgagccgttcgcca 445
           |  |||||| | ||| ||||  | | |  |   |||| || || |    || |     |
Db    6777 CGACGCTGCTCGCCCATCCGGACCGGATGACGACTGCGGTGGACGAGATGCTGCGGTACT 6718

Qy     446 tcttcgtcaatgcgatggacaccaatccgctggctgccgaccctacggtcat 497
           |||  |||||  |       | |||  | ||| ||  ||||    | | ||||
Db    6717 TCTCGGTCAACGACGGAACCCCCATGCGGCTCGCCACCGAGGACATGCTCAT 6666


RESULT  28
AP003402/c
LOCUS       AP003402               187948 bp    DNA     linear   PRI 06-NOV-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-661I21,
            complete sequence.
ACCESSION   AP003402
VERSION     AP003402.2  GI:16751506
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:RP11-661I21.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187948)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Nov 5, 2001 this sequence version replaced gi:13366105.
FEATURES             Location/Qualifiers
     source          1. .187948
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-661I21"
BASE COUNT    55717 a  41857 c  40668 g  49706 t
ORIGIN


  Query Match              2.9%;  Score 39.2;  DB 9;  Length 187948;
  Best Local Similarity   51.7%;  Pred. No. 79;
  Matches   89;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     414 aattcctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatcc 473
           ||||||||||   | ||  ||       |  |    ||    | |   ||| ||  ||||
Db   29076 AATTCCTGCCAAGGGTGGTGAAGTTCCTGGGAGGACCGGTGGTACTGAGGAAACAGATCC 29017

Qy     474 gctggctgccgaccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcct 533
             |||   ||  |  |    | | || | | |  | |  | ||  ||| |  |   || |
Db   29016 TTTGGGACCCAGCTTTTGTCTGAGTAACGAGGGTGACTTCAAGTGTTTTAGGCAAAGCTT 28957

Qy     534 gttggtattgagccgtttgaccgaacgcaaaatccatgtttgtaaggcagct 585
           |||  | || ||  |||| ||| |  | |||||  | |  || ||   ||||
Db   28956 TTTGCTTTTCAGAGGTTTCACCCACTGAAAAATGTAGGGGTGGAATAGAGCT 28905


RESULT  29
SC5F2A
LOCUS       SC5F2A                 37245 bp    DNA     linear   BCT 05-MAY-2000
DEFINITION  Streptomyces coelicolor cosmid 5F2A.
ACCESSION   AL049587
VERSION     AL049587.2  GI:7768282
KEYWORDS    ABC transporter; amino acid permease; BCCT family;
            carboxypeptidase; D-amino acid oxidase; ddah, dimethylarginine
            dimethylaminohydrolase; dehydrogenase; efflux protein;
            endonuclease; enoyl CoA hydratase; IPP isomerase; metallopeptidase;
            oxidoreductase; racemase; regulator; thiolase; transferase.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 37245)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 37245)
  AUTHORS   Oliver,K. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 37245)
  AUTHORS   Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-1999) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     On May 10, 2000 this sequence version replaced gi:4584464.
            Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
```

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/ jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid 5F2A.

FEATURES Location/Qualifiers

source 1. .37245
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 5F2A"

misc_feature 1. .105
/note="Nominal overlap with cosmid 4C6"

gene complement(83. .859)
/gene="SC5F2A.01c"
/note="ddah"

CDS complement(83. .859)
/gene="SC5F2A.01c"
/note="SC5F2A.01c, ddah, dimethylarginine dimethylaminohydrolase, len: 258 aa; hydrolyses asymmetrically methylated arginine with preference for dimethylated arginine over monomethylated arginine (experimental). Similar to TR:D1038106 (EMBL:AB001915) NG,NG-dimethylarginine dimethylaminohydrolase from Homo sapiens (285 aa) fasta scores; opt: 323, z-score: 384.7, E(): 4.1e-14, (35.7% identity in 272 aa overlap)."
/codon_start=1
/transl_table=11
/label=ddah
/product="dimethylarginine dimethylaminohydrolase"
/protein_id="CAB40668.1"
/db_xref="GI:4584465"
/db_xref="SWISS-PROT:Q9X7M4"
/translation="MPSKKALVRRPSPRLAEGLVTHVEREKVDHGLALEQWDAYVEAL
GAHGWETLEVDPADDCPDSVFVEDAVVVFRNVALITRPGAESRRAETAGVEEAVARLG
CSVNWVWEPGTLDGGDVLKIGDTIYVGRGGRTNAAGVQQLRAAFEPLGARVVAVPVSK
VLHLKSAVTALPDGTVIGHIPLTDVPSLFPRFLPVPEESGAHVVLLGGSRLLMAASAP
KTAELLADLGHEPVLVDIGEFEKLEGCVTCLSVRLRELYD"

gene complement(976. .3234)
/gene="SC5F2A.02c"

CDS complement(976. .3234)
/gene="SC5F2A.02c"
/note="SC5F2A.02c, UvrA-like ABC transporter, len: 752aa; similar to many egs. TR:O86699 (EMBL:AL031515) UvrA-like protein from Streptomyces coelicolor (796 aa) fasta scores; opt: 2047, z-score: 2164.0, E(): 0, (51.9% identity in 776 aa overlap) and SW:UVRA_ECOLI UvrA, excinuclease ABC subunit A from Escherichia coli (940 aa) fasta scores; opt: 634, z-score: 669.2, E(): 5.8e-30, (33.9% identity in 938 aa overlap). Contains two Pfam matches to entry PF00005 ABC_tran, ABC transporter and two Prosite matches to PS00017 ATP/GTP-binding site motif A (P-loop). Also contains possible membrane spanning hydrophobic regions."
/codon_start=1
/transl_table=11
/label=SC5F2A.02c
/product="UvrA-like ABC transporter"
/protein_id="CAB40669.1"
/db_xref="GI:4584466"
/db_xref="SPTREMBL:Q9X7M5"
/translation="MSEFISITGARENNLQDVTLRIPKGRLTVFTGVSGSGKSSVVFD
TIAVESRRQLNETFTWFVRNRLPKYERPHADALEGLTPAIVVDQRPVGGHSRSTVGTM
TDIHSVLRVLFSRHGTPGAGGATAYSFNDPSGMCPGCDGLGRRVQPDWDRILDPARSL
ADGAVRFPPFAAGTWQGQTYTNTEELDTGKPVGDFTAAERAFLMRGRPGSKVTVSGSG
GTWSTEYEGLADRFERLYLKRDLSGMSERTRDLVRGFLVEARCPDCGGARLNAAALAS
RIDGHSIADCSRMQITDLIAVLRGIDDPVALPVAGAAVAALERVEAIGLGYLSLDRET
ATLSGGEGQRLKTVRHLGSSLTGMTYIFDEPSVGLHPRDVGRLGDLLLRLRDKGNTVL
VVEHDPDVIALADHVVDMGPRAGADGGRVVFEGTPAGLAASDTLTGRCLGRRTAVKDT
VRAPTGELWVKGAERHNLREVTVAFPTGVLTAVTGVAGSGKSTLVAELTGAHPDAVVV
DQSAIGISARSTPATYLGIMDTVRKVFARETGAEPGFFSFNSAGACGTCEGRGIIHTD
LAFMDPVTTTCHDCEGRRFREEVLRLTVDGRSVADVLAMTAGQALGFFSDPGVRRRLR
ALRDVGLTYLTLGQPLSTLSGGERQRIKLATRLHRTGAVYVLDEPTTGLHMSDVEGLL
ALLDRLVDAGNTVVVVEHNLDVVAHADRVIDLGPDGGRDGGRVIFEGTPRELLAARGS
STAEHLRRATRR"

misc_feature complement(1090. .1803)
/gene="SC5F2A.02c"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 101.40, E-value 1.7e-26."

misc_feature complement(1759. .1782)
/gene="SC5F2A.02c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."

misc_feature complement(2002. .3162)
/gene="SC5F2A.02c"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 35.60, E-value 4e-08."

misc_feature complement(3118. .3141)
/gene="SC5F2A.02c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."

gene complement(3245. .4879)
/gene="SC5F2A.03c"

CDS complement(3245. .4879)
/gene="SC5F2A.03c"
/note="SC5F2A.03c, probable ABC transporter, len: 544aa; similar to many involved in antibiotic resistance egs. TR:Q54381 (EMBL:X79146) from the lincomycin-production gene cluster of Streptomyces lincolnensis 78-11 (578 aa) fasta scores; opt: 1624, z-score: 1696.3, E(): 0, (50.3% identity in 541 aa overlap). Contains two Pfam matches to entry PF00005 ABC_tran, ABC transporter, two Prosite matches to PS00017 ATP/GTP-binding site motif A (P-loop) and Prosite match to PS00211 ABC transporters family signature. Also contains a possible coiled-coil region around 250aa."
/codon_start=1
/transl_table=11
/label=SC5F2A.03c
/product="putative ABC transporter"
/protein_id="CAB40670.1"
/db_xref="GI:4584467"
/db_xref="SPTREMBL:Q9X7M6"
/translation="MSTSPTSLTCTSLSFAWPDGTEVFDGLDVAFGPGRTGLVGLNGA
GKSTLLKLIAGRLTPADGTVRVAGQVGYLPQNVTLDTALRVDEALGIDGRRAALHAIE
AGDVREEHFETVGDDWDVEERALATLGELGLDHIGLDRTVGEVSGGESVLLRLAALLL
RRPDVLLLDEPTNNLDVYARRRLYAAVESWPGVMVVVSHDRELLDRVDQIADLRAGSV
TWYGGNLTAYEEALAVEQEAAERMVRVAESDLRRQKRELADAQVVLARRRRYGQKMYD
TKREPRAVMKLRARTAQQSAGKYRIMHEEKLAGAKERLDDAVEAVRDDDEIRVDLPYT
AVPPGRTVLTLRALELAYGARVAGGLDLHGPERIALLGRNGAGKTTLLRTVAGELAPV
AGEATAHVPLRFLPQRLDVLDDGLTVAENVARFAPGATNNRIRALLARFLFRGARADQ
QAATLSGGERFRAALAALMLAEPAPQLLMLDEPTNNLDMASVRQLTGALEAYEGALLV
ASHDLPFLESLGITRWLLLEEGELREITPEAVGYSA"

misc_feature complement(3290. .3787)
/gene="SC5F2A.03c"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 122.30, E-value 9e-33"

misc_feature complement(3743. .3766)
/gene="SC5F2A.03c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."

misc_feature complement(4232. .4783)
/gene="SC5F2A.03c"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 143.80, E-value 3.1e-39"

misc_feature complement(4403. .4447)
/gene="SC5F2A.03c"
/note="PS00211 ABC transporters family signature."

```
     misc_feature    complement(4739. .4762)
                     /gene="SC5F2A.03c"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)."
     gene            4942. .5850
                     /gene="SC5F2A.04"
     CDS             4942. .5850

  Query Match             2.8%;  Score 38.2;  DB 1;  Length 37245;
  Best Local Similarity   44.4%;  Pred. No. 1.3e+02;
  Matches  154;  Conservative    0;  Mismatches  193;  Indels    0;  Gaps    0;

Qy     967 cagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcgccgcag 1026
           ||||    ||   |||||    ||||| |    | | |  || |  | || | | |
Db   23451 CAGAAGGGCGCGCTCGAACCCAGCCGCTGGGTGGTGGTCGTCTGGGGTGTGGTGACCGGC 23510

Qy    1027 ccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaactcttc 1086
           ||| |     |  |||   ||  |    |  |||  |             ||    |
Db   23511 GCGGTCGCGGCCATCATGCTGCTGATCGGCAGCGGTGAGGCGGACGCGCTCACCGGTCTG 23570

Qy    1087 aagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtacttacgag 1146
           ||  |  |||  | ||   |  |    |  || ||| |||  |||    |  | ||
Db   23571 CAGAACCTCACGATCCTCGCGGCGGCGCCCTTCGTCATCGTGATGATCTTCATGTCGTC 23630

Qy    1147 cgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtcggcgat 1206
             | | |||| |        ||    | | ||  |  ||||||     || | |. |||
Db   23631 GCCCTCATGCGCGACCTGCGCCGCGACCCGCTCATGGTGCGCGGCGAGATGGCCAACGAA 23690

Qy    1207 accgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgctttgtgc 1266
           |||  |     |   | | |   ||     |   | |||| |  ||| |||   |  |
Db   23691 GCCGTCGAACTGGCCGTCATCGAGGGCCACAAGCAGTACGAGGGCGACTTCGAGATCCGG 23750

Qy    1267 agcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaa 1313
           | |  |  | |||| |||| | |||   | |||   | | |  ||||
Db   23751 ATCGGCCCCGGCCCCGGCACAGACGTGGAGGGCGACCCGATCGGCAA 23797


RESULT  30
SC10A5
LOCUS       SC10A5                  38734 bp    DNA     linear   BCT 26-APR-2000
DEFINITION  Streptomyces coelicolor cosmid 10A5.
ACCESSION   AL021529
VERSION     AL021529.1  GI:2815309
KEYWORDS    ABC-transporter; cold-shock domain protein; DEAD-box RNA helicase;
            NTP pyrophosphohydrolase; oxidase; oxidoreductase; periplasmic
            oligopeptide binding protein; phosphoenolpyruvate-utilizing enzyme;
            secreted protease; transcriptional regulator; transposase
            pseudogene; xlnA; xylanase A.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 38734)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 38734)
  AUTHORS   Murphy,L. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 38734)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
            program of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid 10A5 lies between St3F7 and St7H1 on the AseI-B genomic
            restriction fragment.
FEATURES             Location/Qualifiers
     source          1. .38734
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 10A5"
     gene            1. .289
                     /gene="SC10A5.01"
     misc_feature    1. .108
                     /gene="SC10A5.01"
                     /note="overlap with St3F7 from: 23623 to: 23730"
     CDS             <1. .289
                     /gene="SC10A5.01"
                     /note="SC10A5.01, probable phosphoenolpyruvate-utilizing
                     en zyme, partial CDS, len: >95 aa; overlaps and extends
                     ORF fr om neighbouring cosmid SC3F7.16; similar to many
                     eg. PPSA_S TAMA P46893 probable phosphoenolpyruvate
                     synthase (834 aa), fasta scores; opt: 215 z-score: 351.4
                     E(): 2.3e-12, 47.8% identity in 69 aa overlap"
                     /codon_start=2
                     /transl_table=11
                     /label=SC10A5.01
                     /protein_id="CAA16433.1"
                     /db_xref="GI:2815310"
                     /db_xref="SPTREMBL:O54094"
                     /translation="ILVTGYTDASWTPLFAIAGGVVTDIGSMLSHSSIVAREFHVPSV
                     VNTKDATQRINTGDLIVVDGDAGTVEVVESADTDPQGPAGAAGTPAGATTD"
     RBS             355. .358
                     /note="possible RBS upstream of SC10A5.02"
     gene            367. .1554
                     /gene="SC10A5.02"
     CDS             367. .1554
                     /gene="SC10A5.02"
                     /note="SC10A5.02, probable oxidase, len: 395; similar in
                     N- terminus to many diverse oxidases eg. TR:P95483
                     (EMBL:U7449 3) AMINOPYRROLNITRIN OXIDASE PRND from
                     PSEUDOMONAS FLUORESC ENS (363 aa), fasta scores; opt: 349
                     z-score: 372.1 E(): 1. 6e-13, 33.3% identity in 165 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /label=SC10A5.02
                     /protein_id="CAA16434.1"
                     /db_xref="GI:2815311"
```

```
                     /db_xref="SPTREMBL:O54095"
                     /translation="MIPNQWYPIVEAQEVGNDKPLGVRRMGQDLVLWRDIDGNLVCQG
                     ARCPHKGANLGDGRMKGNTIECPYHGFRYGADGACRVIPAMGSEARIPGSLRVPTYPV
                     REQFGLVWMWWGDERPTADLPPVAAPAEVTDNRKLYATKRWTRPVHYTRYIESLLEFY
                     HVTYVHRDHWFNYIDYLLLYGTPSKFGLDGRERYLAATRITNHRVETEAEGQTIRYSF
                     DHCQEDDPTNTTHYVITFTFPCMVHVQTEQFETTSWLVPIDDQNTEHILRWYEYEQVK
                     PVLRFEPLRRLLPWASLYMEKWVQDPQDVRIMEHQEPKISAGGVNKFIPVDEMNAKYI
                     SMRAKLIADASAAPSSPARAAEPEPEAAGRGGSAARATGNGRGAAGGRRGTKPKEDAA
                     ARP"
gene                 1597. .2277
                     /gene="SC10A5.03"
CDS                  1597. .2277
                     /gene="SC10A5.03"
                     /note="SC10A5.03, probable membrane protein, len: 226 aa"
                     /codon_start=1
                     /transl_table=11
                     /label=SC10A5.03
                     /protein_id="CAA16435.1"
                     /db_xref="GI:2815312"
                     /db_xref="SPTREMBL:O54096"
                     /translation="MYGGYDASTGPKALVTAFNTVAVAGAVWFLFGGADTVADWFGTD
                     FDEAVTLRRVLLATLSVLYLLRFIATNFVMLQRKMEWSESATIGIWVLVIHGTMAYFG
                     GTNDAGVSAFTWLGVVLYLLGSYLNTGSEYQRKLWKKRPENKGKLYTEGLFKHSMHIN
                     YFGDAVLFSGFALVTGTPWAFAIPLIMVCMFVFLNIPMLDKYLAERYGEAFDEYASRT
                     AKFVPYVY"
RBS                  2433. .2436
                     /note="possible RBS upstream of SC10A5.04"
gene                 2442. .3329
                     /gene="SC10A5.04"
CDS                  2442. .3329
                     /gene="SC10A5.04"
                     /note="SC10A5.04, unknown, len: 295 aa; similar to
                     hypothet ical proteins from Mycobacterioum tuberculosis
                     and M. lepra e eg. YV29_MYCTU Q11167 hypothetical 39.3 kd
                     protein cy20g9 .29 (358 aa), fasta scores; opt: 1075
                     z-score: 1514.7 E(): 0, 57.8% identity in 263 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /label=SC10A5.04
                     /protein_id="CAA16436.1"
                     /db_xref="GI:2815313"
                     /db_xref="SPTREMBL:O54097"
                     /translation="MTGDIPFGEAEASLTAEVLREVLAGGAEAFARLTSDEGAVDDFG
                     FDPELTDDYLLPALRLLYEKYFRVDLEGLENVPAEGGALLVANHSGTLPLDALMLQVA
                     LHDHHSTHRRLRLLAADLAFDLPVVRDLARKAGHVRACPENALRLLGSGELVGVMPEG
                     YKGLGKPFEERYRLQRFGRGGFAAVALRSRRPMVPCSIVGAEEIYPMIGSAPTLARML
                     KLPYFPITPTFPLLGALGLIPMPTKWTIRFGAPIHTDGFPEDAAEDPLVVEKLAGEVK
                     DTIQHTLNEMLEGRGSPFV"
gene                 3698. .5773
                     /gene="SC10A5.05"
CDS                  3698. .5773
                     /gene="SC10A5.05"
                     /note="SC10A5.05, probable membrane protein, len: 691;
                     high ly similar to M. tuberculosis putative membrane
                     protein TR: O07192 (EMBL:Z96072) MTCY05A6.11C (647 aa),
                     fasta scores; o pt: 1980 z-score: 3557.6 E(): 0, 62.7%
                     identity in 659 aa o verlap and weakly similar to many
                     transport proteins eg. PR OY_SALTY P37460 proline-specific
                     permease proy from Salmone lla typhimurium (456 aa), fasta
                     scores; opt: 169 z-score: 1 82.8 E(): 0.0055, 24.2%
                     identity in 355 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /label=SC10A5.05
                     /protein_id="CAA16437.1"
                     /db_xref="GI:2815314"
                     /db_xref="SPTREMBL:O54098"
                     /translation="MSKLTDVPKRILIGRALRSDRLGETLLPKRIALPVFASDPLSSV
                     AYAPGEVLLVLSIAGVSAYHFSPWIAVAVVVLMFTVVASYRQNVHAYPSGGGDYEVAT
                     TNLGPKAGLTVASALLVDYVLTVAVSISSGIENLGSAIPFVVEHKVLCAVAVILLLTL
                     MNLRGVRESGTLFAIPTYVFVAGVFIMIVWGAFRGLVLDDTMRAPTADYEIKPEHGGL
                     AGFALIFLLLRAFSSGCAALTGVEAISNGVPAFRKPKSKNAGNTLAMMGLLAVTMFCG
                     IIALAAATDVRMSENPATDLFHNGVAVGADYVQHPVISQVAEAVFGEGSFLFIVLAAA
                     TALVLFLAANTAYNGFPLLGSILAQDRYLPRQLHTRGDRLAFSNGIVLLAGAAMLLVV
                     VYGADSTRLIQLYIVGVFVSFTLSQIGMVRHWNRNLAGERDQSKRRHMMRSRAINAFG
                     AFFTGLVLVVVLATKFTHGAWVALLGMCIFFATMTAIRKHYDRVAEEIAAPEDPEEAQ
                     SDDMVRPSRVHSVVLISKIHRPTLRALAYAKLMRSDSLEALSVNVDPAETKALREEWE
                     RRGIAVPLKVLDSPYREITRPVIEYVKSLRKESPRDAVSVIIPEYVVGHWYEHLLHNQ
                     SALRLKGRLLFTPGVMVTSVPYQLESSEAARRRARKRQDWSAPGAVRRGPAHHHQDRD
                     RTKDSSSST"
RBS                  5812. .5815
                     /note="possible RBS upstream of SC10A5.06"
gene                 5826. .7202
                     /gene="SC10A5.06"
CDS                  5826. .7202
                     /gene="SC10A5.06"
                     /note="SC10A5.06, unknown, len: 458; similar to M.
                     tubercul osis hypothetical protein TR:O07191 (EMBL:Z96072)
                     MTCY05A6. 10C (405 aa), fasta scores; opt: 455 z-score:
                     878.5 E(): 0, 38.4% identity in 417 aa overlap, and weakly
                     simialar to s everal putative RNA methyltransferases eg.
                     YGCA_ECOLI P5513 5 hypothetical rna methyltransferase (433
```

```
  Query Match              2.8%;  Score 38.2;  DB 1;  Length 38734;
  Best Local Similarity   48.0%;  Pred. No. 1.3e+02;
  Matches  109;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy   1116 ccgcgccatggtgccgattggtacttacgagcgcgtgatgcccttggatatcctgcccac 1175
          | ||| |||  |  || | |   |          ||||  |   | ||| | || | ||||
Db   4519 CTGCGGCATCATCGCGCTGGCCGCCGCGACCGACGTGCGGATGTCGGAGAACCCGGCCAC 4578

Qy   1176 cctgcttttgcgcgatttaatcgtcggcgataccgacagcgcgcaggcattgggttgctt 1235
          |   || || | | |     |||  | ||   | |||  ||  |||     ||    ||
Db   4579 CGACCTCTTCCACAACGGCGTCGCGGTCGGCGCGGACTACGTCCAGCACCCGGTGATCTC 4638

Qy   1236 ggaattggacgaagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaata 1295
          | |  | | ||| |  | | |||    | |||||||| | | |   |       || |
Db   4639 GCAGGTCGCCGAGGCGGTCTTCGGCGAGGGCAGCTTCCTGTTCATCGTGCTGGCCGCAGC 4698

Qy   1296 cggcccgctgttgcgcaaagtgctggaaaccattgagaaggaaggct 1342
          |  | |||| | | |    |    |  | ||  | | |  | ||||
Db   4699 CACCGCGCTGGTCCTCTTCCTCGCCGCCAACACCGCGTACAACGGCT 4745
```

```
RESULT  31
AF103943
LOCUS       AF103943                975 bp    DNA     linear   BCT 15-OCT-1999
DEFINITION  Streptomyces griseus factor C protein precursor, gene, complete
            cds.
ACCESSION   AF103943
VERSION     AF103943.1  GI:4038406
KEYWORDS    .
SOURCE      Streptomyces griseus.
  ORGANISM  Streptomyces griseus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 975)
  AUTHORS   Biro,S., Bekesi,I., Vitalis,S. and Szabo,G.
  TITLE     A substance effecting differentiation in Streptomyces griseus.
            Purification and properties
  JOURNAL   Eur. J. Biochem. 103 (2), 359-363 (1980)
  MEDLINE   80156796
REFERENCE   2  (bases 1 to 975)
  AUTHORS   Birko,Z., Sumegi,A., Vinnai,A., van Wezel,G., Szeszak,F.,
            Vitalis,S., Szabo,P.T., Kele,Z., Janaky,T. and Biro,S.
  TITLE     Characterization of the gene for factor C, an extracellular signal
            protein involved in morphological differentiation of Streptomyces
            griseus
  JOURNAL   Microbiology 145 (9), 2245-2253 (1999)
  MEDLINE   99445165
   PUBMED   10517577
REFERENCE   3  (bases 1 to 975)
  AUTHORS   Birko,Z., Sumegi,A., Szeszak,F., Vitalis,S., Szabo,G. and Biro,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1998) Human Genetics, University Medical School
            of Debrecen, Hungary, Nagyerdei krt 98, Debrecen H-4012, Hungary
```

```
FEATURES             Location/Qualifiers
     source          1. .975
                     /organism="Streptomyces griseus"
                     /strain="45H"
                     /db_xref="taxon:1911"
     sig_peptide     1. .114
     CDS             1. .975
                     /function="involved in the regulation of
                     cytodifferentiation"
                     /codon_start=1
                     /transl_table=11
                     /product="factor C protein precursor"
                     /protein_id="AAC97368.1"
                     /db_xref="GI:4038407"
                     /translation="MPAATTSSPSRRGLLRLGGGLVAASAFGLSFGAGSAAAAVPATK
                     RFSLTEPSHDLFRHAKLHDARVQQSFTFDIVNRRLFVAQLKSGSPDDSGDLCITQLDF
                     SGNKLGHMYLLGFGHGVSIGAQPVGADTYLWTEVDVNSNARGTRLARFKWNNGATLSR
                     TSSALAKHQPVPGATEMTCAIDPVNNRMAIRYLTASGRRYGIYNVADIAAGVYDKPLS
                     DVPHPTGLGTFQGYALYGSYVYQLTGNPYGPDNPNPGNSYVSSVDVNTGALVQRAFTR
                     AGSTLTFREPEGMGIYRTAAGEVRLFLGFASGVAGDRRSNLFYKNVLI"
     mat_peptide     115. .972
                     /product="factor C protein"
BASE COUNT      149 a    382 c    303 g    141 t
ORIGIN
```

```
  Query Match             2.8%;  Score 38;  DB 1;  Length 975;
  Best Local Similarity   45.8%;  Pred. No. 1.2e+02;
  Matches  131;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy   1026 gccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaaaacaaactctt 1085
          || || | |  | ||  |||| | | ||  ||     |||          |  ||
Db    192 GCGGGTCCAGCAGTCGTTCACCTTCGACATCGTCAACCGGCGGCTGTTCGTGGCCCAGTT 251

Qy   1086 caagttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtacttacga 1145
           ||||    | |  ||| | || ||||||||||    |||    | | | |     |  |
Db    252 GAAGTCGGGCTCGCCCGACGACAGCGGCGACCTGTGCATCACCCAGCTGGACTTCTCGGG 311

Qy   1146 gcgcgtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtcggcga 1205
             | | ||  |        |||| |  | |        |||  |  |||| || | |
Db    312 CAACAAGCTGGGCCACATGTACCTGCTCGGCTTCGGCCACGGCGTCTCCATCGGCGCCCA 371

Qy   1206 taccgacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgctttgtg 1265
            | | | ||||| |  | |     ||   |||  ||||||   |   |  |||     |
Db    372 GCCGGTCGGCGCGGACACGTACCTGTGGACGGAGGTGGACGTCAACAGCAACGCGCGTGG 431

Qy   1266 cagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgc 1311
           || |    |   || |  ||| |   |  ||||| ||  | |  ||
Db    432 CACCCGGCTGGCCCGGTTCAAGTGGAACAACGGCGCGACGCTCAGC 477
```

```
RESULT  32
ATY14325
LOCUS       ATY14325                1694 bp    mRNA    linear   PLN 25-JUN-1999
DEFINITION  Arabidopsis thaliana mRNA for mevalonate diphosphate decarboxylase.
ACCESSION   Y14325
VERSION     Y14325.1  GI:2288886
KEYWORDS    mevalonate diphosphate decarboxylase; MVD1 gene.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1694)
  AUTHORS   Cordier,H., Karst,F. and Berges,T.
  TITLE     Heterologous expression in Saccharomyces cerevisiae of an
            Arabidopsis thaliana cDNA encoding mevalonate diphosphate
            decarboxylase
  JOURNAL   Plant Mol. Biol. 39 (5), 953-967 (1999)
  MEDLINE   99273998
REFERENCE   2  (bases 1 to 1694)
  AUTHORS   Cordier,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1997) H. Cordier, IBMIG, 40, avenue du Recteur
            Pineau, 86022 Poitiers Cedex, FRANCE
COMMENT     Related sequence: T44736.
FEATURES             Location/Qualifiers
     source          1. .1694
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     gene            118. .1356
                     /gene="MVD1"
     CDS             118. .1356
                     /gene="MVD1"
                     /EC_number="4.1.1.33"
                     /codon_start=1
                     /product="mevalonate diphosphate decarboxylase"
                     /protein_id="CAA74700.1"
                     /db_xref="GI:2288887"
                     /db_xref="SPTREMBL:O23722"
                     /translation="MAEEKWVVMVTAQTPTNIAVIKYWGKRDEVRILPINDSISVTLD
                     PDHLCTLTTVAVSPSFDRDRMWLNGKEISLSGSRYQNCLREIRSRADDVEDKEKGIKI
                     AKKDWEKLHLHIASHNNFPTAAGLASSAAGFACLVFALAKLMNVNEDPSQLSAIARQG
                     SGSACRSLFGGFVKWNMGNKEDGSDSVAVQLVDDKHWDDLVIIIAVVSSRQKETSSTS
                     GMRESVETSLLLQHRAKEVVPVRILQMEEAIKNRDFTSFTKLTCSDSNQFHAVCMDTS
                     PPIFYMNDTSHRIISLVEKWNRSAGTPEIAYTFDAGPNAVMIARNRKVAVELLQGLLY
                     CFPPKPDTDMKSYVLGDTSIVKEAGLEGELPQGIKDKIGSQDQKGEVSYFICSRPGRG
                     PVVLQDQTQALLHPQTGLPK"
BASE COUNT      480 a    352 c    375 g    487 t
ORIGIN
```

```
  Query Match             2.8%;  Score 37.8;  DB 8;  Length 1694;
  Best Local Similarity   55.8%;  Pred. No. 1.4e+02;
  Matches   72;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy    128 tgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagacaaaaaga 187
          | ||||  |   |   |  ||  || ||||||    ||   ||||    |   |  |||
Db     71 TCAAAGCTACACAGATCTCTGGAATCGAAAAAGCTTGAGCTGTGTTAATGGCGGAGGAGA 130

Qy    188 atccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcaccgtggcg 247
          |   ||  ||| || | || |||| | || ||   || |||||||  |||| |   | | |
Db    131 AATGGGTGGTGATGGTGACGGCGCAGACTCCAACGAATATCGCCGTGATTAAGTATTGGG 190

Qy    248 aaaagcgcg 256
           |||| | |
Db    191 GAAAGAGAG 199
```

```
RESULT  33
AE004619
LOCUS       AE004619               11351 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 180 of 529 of the complete
            genome.
ACCESSION   AE004619 AE004091
VERSION     AE004619.1  GI:9947920
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 11351)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 11351)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
```

```
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .11351
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            281. .2131
                     /gene="PA1926"
     CDS             281. .2131
                     /gene="PA1926"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG05314.1"
                     /db_xref="GI:9947921"
                     /translation="MPGTDAAGCTRIPGKDLPLEQTIANLSAILAGLGMKIEIASWRN
                     IVPNVWSLHIRDAQSPMCFTNGKGASKESALASALGEFIERLNCNFFYNDQYWGEEIA
                     NAAFVHYPDERWFKPGRDDALPLGLLDGHCLAIYDPDGELRGSHLYDTNSGNVQRGIC
                     ALPFVRQSDGQVVYFPSNLIENLYLSNGMSAGNTLAEAQVQCLSEIFERAVKRQILEG
                     ELALPDVPPEVLAKYPGILAGIRGLEEQGFPVLVKDASLGGEFPVMCVTLMNPRTGGV
                     FASFGAHPSFEVALERSLTELLQGRSFEGLNDLPQPTFESHALTEPNNFVEHFIDSSG
                     VVSWRFFSAKADFPFVEWDFTRQGEAANAEEAATLFGILEAMGKQVYMAVYEHLGATA
                     CRILVPGYSEIYPVEDLIWDNTNKALAFREDILNLHRLDDAALGALLERLEDCEVDDY
                     TDITTLIGVEFDDNTVWGQLTILELKVLIGLALKRFEDAKEGVEAFLQYNDNSVERGL
                     FYQALNVVLEVVLDDELEIADYEANFRRMFGDARMDAALGSVDGSVRFFGLTPTSMKL
                     EGLDRHLRLIDSYKKLHAARARMQPVVDGEAGAAAAGGLKPRRMAIRKRK"
     gene            2264. .4564
                     /gene="metE"
                     /note="PA1927"
     CDS             2264. .4564
                     /gene="metE"
                     /codon_start=1
                     /transl_table=11
                     /product="5-methyltetrahydropteroyltriglutamate-
                     homocysteine S-methyltransferase"
                     /protein_id="AAG05315.1"
                     /db_xref="GI:9947922"
                     /translation="MALAHTLGFPRIGRDRELKKAQEAFWKGELDEAGLRAVGRQLRA
                     AHWQVQKDAGIQLLPVGDFAWYDQVLTHSLTFGVIPERFRAHGEAGPTLHTLFGMARG
                     VSDDSCCGGAHAQEMTKWFDTNYHYLVPEFSADQQFQLSWEQLFEEVDEARALGHQVK
                     PVLIGPLTYLWLGKAKGAEFDRLDLLDRLLPLYGQIFRRLAGQGVEWVQIDEPILVLD
                     LPQAWKNAFERAYNLLQSEPLKKLVATYFGGLEDNLGLAANLPVDGLHIDLVRAPEQY
                     PSILDRLPAYKVVSLGLVNGRNVWRCDLEKALEVVRHARERLGERLWVAPSCSLLHSP
                     VDLEREDGLDAELKSWLAFAVQKCREVAVLARAATEPEAAEVLAALEESRAVQASRAS
                     SPRIHKPAVQARLAAIKASDAQRRSPFAERIARQRAGLDLPAFPTTTIGSFPQTSSIR
                     LARQSFKQGKLSEAEYIEAMHSEIRHAVQIQEQLGLDVLVHGEAERNDMVEYFAEQLD
                     GYVFTRFGWVQSYGSRCVKPAVIYGDLSRPRAMTVEWIRYAQSLTDKVMKGMLTGPVT
                     MLMWSFPREDVSREVQARQLALAIRDEVVDLEAAGIRIVQIDEAAFREGLPLRRNAWP
                     HYLEWATEAFRLCASGVRDETQIHTHMCYSEFNDVIESIAAMDADVITIETSRSDMEL
                     LEAFEQFDYPNEIGPGVYDIHSPRVPSREEIVALLRKAARRIPAERLWVNPDCGLKTR
                     AWPETEAALVNMVAAARELRGDLARG"
     gene            complement(4626. .5195)
                     /gene="rimJ"
                     /note="PA1928"
     CDS             complement(4626. .5195)
                     /gene="rimJ"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosomal protein alanine acetyltransferase"
                     /protein_id="AAG05316.1"
                     /db_xref="GI:9947923"
                     /translation="MPGNRRQVPILHRDADFLLRALQPDDFVRTSRYENANREHLAPW
                     EPLRDPGYFSVDNARARTLLQVASMDEGEALLLLLLDPDDGEVLGRCSYTNIVRGVFQ
                     ACHLGFSLAAAAQGRGLMARALRVANRYCFEQLGLHRIMASHLPRNARSERLLESLGF
                     EKEGYARAYLKIAGVWEDHVLRALVDAPR"
     gene            complement(5242. .5538)
                     /gene="PA1929"
     CDS             complement(5242. .5538)
                     /gene="PA1929"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG05317.1"
                     /db_xref="GI:9947924"
                     /translation="MATDLYALAEQVEDEAGFLRFLAALAADWEEDRRLAAANPPKPF
                     GPSPLGWENGSIGGFLDAAATWGEDSRDGLPGYSRPENAWRRAAQLMLAGKFYE"
     gene            5910. .7205
                     /gene="PA1930"
     CDS             5910. .7205
                     /gene="PA1930"
                     /codon_start=1
                     /transl_table=11
                     /product="probable chemotaxis transducer"
                     /protein_id="AAG05318.1"
                     /db_xref="GI:9947925"
                     /translation="MLFGRKSRAVESAAKDEDLAWRLDAIGQNVATIRFTPDGQILSA
                     NPLFLAVVGYSADELVGKHHRIFCEEDFQASAAYVRFWKELASGTPQRGVFKRLRRDG
                     TPVWLEATYFPVKNAEGAVVEVLKIAADVTRNHSELLLLNAINDAIRQSMAVIEFTPD
                     GEILDANENFLRLFGYSLKSLKGQHHRMLCFDEFYRENPDFWARLRHGEFSRGHFERR
                     SAAGERVHIEATYNPVKDSSGRIIKVIKFAIDVTEQVNRNEEVRRAAELSFSTAEETA
                     QISTRGIDSLDQSVALSAKTLSMVNTAVGLISQLTHQTKDIENIVTTIQSVAEQTNLL
                     ALNAAIEAARAGDMGRGFAVVADEVRQLAARTSGATVEIKNVVTENVKLTQQINHNMQ
                     TIESSAQENNNQITTVAGIMQEISKGAEHVARVVSSLFK"
     gene            7375. .7887
                     /gene="PA1931"
     CDS             7375. .7887
                     /gene="PA1931"
                     /codon_start=1
                     /transl_table=11
                     /product="probable ferredoxin"
                     /protein_id="AAG05319.1"
                     /db_xref="GI:9947926"
                     /translation="MTNPEPADLPISLLINGQRRELRVPPWLTLLDLLRERLDLVGSK
                     KGCDHGQCGACTVLRNGRRINACLTLAVMCDGDALTTIEGLAEGDRLHPLQQAFIRHD
                     AFQCGYCTPGQICSALGLASEGRARSRAEIREGMSGNLCRCGAYGNILDAVEDALPLL
                     DRDGQGGQRA"
     gene            7884. .8873
                     /gene="PA1932"
     CDS             7884. .8873
                     /gene="PA1932"
                     /codon_start=1
                     /transl_table=11
                     /product="probable hydroxylase molybdopterin-containing
                     subunit"
                     /protein_id="AAG05320.1"
                     /db_xref="GI:9947927"
                     /translation="MTPFAYAKPADIEQALGLAGPESRYIAGGTNLLDLMKENVARPR
                     RLIDITGLPLRDIRRTPTGGLLVGALVSNAELAWNPAVEADYPLLAQAILAGASPQLR
                     NMASTGGNLLQRTRCQYFYDTGTPCNKREPGSGCPARSGLNRNHAILGASEHCVAVHP
                     SDLCVALAALEAQVLVRGRAGERRVAFADFHRLPADTPERDNCLAVDELVTAIELPAP
                     RFARHCHYLKVRDRASYAFALVSVAAALEPDGERIGELRLALGGVAHKPWRVAELEER
                     FRGHPANRDSYVRFAEALLRDARPLAHNAFKIGLARNAIVRALEEAHSGGIGQ"
     gene            8870. .11074
                     /gene="PA1933"
     CDS             8870. .11074
                     /gene="PA1933"
                     /codon_start=1
                     /transl_table=11
                     /product="probable hydroxylase large subunit"
                     /protein_id="AAG05321.1"
                     /db_xref="GI:9947928"
                     /translation="MSSLPHSTGQALDRVDGPAKVTGQARYAAEYPAADLLHGSVVSS
                     DVARGRVRGIDCAAALAVPGVVAVLHHLNRPPMAGDDEPYKDADAAEGEPFRPLFDDR
                     VLYSGQPLALVLARSLELARYAGSLLRIDIEPEPHQTDLLAALDQAHEAPAQLPAERG
                     DFYTAYGAAPIRVEASYSTPIEHHNPMEPHASTVIVQPDGSLLVHDKTQGTQNSQAYL
                     QKVFGLPADKVRVCAAYVGGAFGSGLRPQYQLALAVMAALQLRRSVRVVLTRQQMFTF
                     GYRPRTLQRLQLGADAEGRLLALGHQATAQTSRFEDFTEHVVEWSGMLYRCENFSLAY
                     RLVPLDVYTPLDMRAPGAALGLIGLECAMDELAVRLGMDPIALRRFNFAERNGNEDKP
                     YSSKALLACYEEGARRFGWQARDPRPRSMSEGRQLLGWGMAGGVWEAMQSKASARARL
                     EADGLLRVASATTDIGTGTYTVMTQIAADAAGMAPEEVRFQLGDSALPKAPLQGGSFT
```

```
                     VSSVGSAVRLACLRLRQALVAHARARHPELAGTAEERFRICRGHLEAGDTRYALADLL
                     QGLPEDALQVEVEAEPSARRKGYATATHSAVFVEVRIDEALGTLRVSRVVSAVAAGRV
                     INPKTARSQILGGVVWGLGMALQEETQVDHALGRCMNHNLAEYHIPVNADIGDIEVIF
                     VDEPDDIVNELGSKGVGEIGIVGVAAAVANAVYHATGRRLRDFPLTVDKLLG"
BASE COUNT     1953 a   3872 c   3777 g   1749 t
ORIGIN


  Query Match             2.8%;  Score 37.8;  DB 1;  Length 11351;
  Best Local Similarity   51.5%;  Pred. No. 1.5e+02;
  Matches   87;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy   1055 ccctcggccatttcctgaaaaacaaactcttcaagttcaacacagccgtcaacggcggcg 1114
          | || ||||  | | ||||  |||| ||| | ||| | |||   | || |||| ||
Db  10833 CGCTGGGCCGCTGCATGAACCACAACCTCGCCGAGTACCACATCCCGGTTAACGCCGATA 10892

Qy   1115 accgcgccatggtgccgattggtacttacgagcgcgtgatgcccttggatatcctgccca 1174
             |||  || | |  |||      | |||| | ||       | |  |    |||  |
Db  10893 TTGGCGATATCGAGGTGATCTTCGTTGACGAACCCGACGATATCGTTAACGAACTGGGCT 10952

Qy   1175 ccctgcttttgcgcgatttaatcgtcggcgataccgacagcgcgcaggc 1223
          ||  |    |  ||||  ||  | ||| ||    || |  |||   |||
Db  10953 CCAAGGGCGTCGGCGAAATAGGCATCGTCGGCGTCGCCGCCGCAGTGGC 11001


RESULT  34
AE004956/c
LOCUS       AE004956               11660 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 517 of 529 of the complete
            genome.
ACCESSION   AE004956 AE004091
VERSION     AE004956.1  GI:9951760
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 11660)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 11660)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .11660
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            95. .544
                     /gene="PA5432"
     CDS             95. .544
                     /gene="PA5432"
                     /codon_start=1
                     /transl_table=11
                     /product="probable acetyltransferase"
                     /protein_id="AAG08817.1"
                     /db_xref="GI:9951761"
                     /translation="MSTSLQIRPVGAGDHAAWLPLWEGYQRFYKTEIAAATSDATWQR
                     FLDADEPMNAALAWLDERAVGLVHWIYHRSCWTVGDYCYLQDLFVAEGLRGEGIGRRL
                     IEHVYAEAGAAGCSRVHWLTHESNSDAMKLYERIADKSGFLQYRRIL"
     gene            553. .1227
                     /gene="PA5433"
     CDS             553. .1227
                     /gene="PA5433"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG08818.1"
                     /db_xref="GI:9951762"
                     /translation="MREPLNWRPVATPPHAPLDGRFVKLEPLDPQRHGDDLWDALQGP
                     GADPLLWEYLAYGPFAERPAFDAWLAGQAASRDPLFFAVRERRTQRAVGLLSYLRITP
                     KDGVIEIGHIAFGAPMQRTPGSTEAVYLLAELALGRLGYRRLEWKCNAGNARSLRAAE
                     RLGFTYEGTFRQHMVVKQRNRDTAWFSLLDSEWPARRAAFERWLDPSNFDAEGRQRQT
                     LEALRD"
     gene            complement(1252. .2505)
                     /gene="mtr"
                     /note="PA5434"
     CDS             complement(1252. .2505)
                     /gene="mtr"
                     /codon_start=1
                     /transl_table=11
                     /product="tryptophan permease"
                     /protein_id="AAG08819.1"
                     /db_xref="GI:9951763"
                     /translation="MSSSPAQTPSRRPSLLGGSMIIAGTAVGAGMFSLPIAMSGIWFG
                     WSVAVFLLTWFCMLLSGMMILEANLNYPVGSSFSTITRDLLGQGWNVVNGLSIAFVLY
                     ILTYAYISGGGSIIGYTLSSGLGVTLPEKLAGLLFALAVALVVWWSTRAVDRITTLML
                     GGMIITFGLSISGLLGRIQPAILFNSGEPDAVYWPYLLATLPFCLTSFGYHGNVPSLM
                     KYYGKDPQRISRSLWIGTLIALAIYLLWQASTLGTIPREQFKGIIAGGSNVGTLVEYL
                     HRITASDSLKALLTTFSNLAVASSFLGVTLGLFDYLADLCRFDDSHFGRFKTALLTFV
                     PPTIGGLLFPNGFIYAIGFAGLAAAFWAVIVPALMARASRKRFGSPLFRAWGGTPAIV
                     LVLLFGVANAVAHILASLHWLPEYR"
     gene            complement(2686. .4509)
                     /gene="PA5435"
     CDS             complement(2686. .4509)
                     /gene="PA5435"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transcarboxylase subunit"
                     /protein_id="AAG08820.1"
                     /db_xref="GI:9951764"
                     /translation="MSNTIQAKKTGVTDTILRDAHQSLLATRMRTEDMLPICDKLDRV
                     GYWSLEVWGGATFDACVRFLKEDPWERLRKLKAALPNTRLQMLLRGQNLLGYRHYSDD
                     VVRAFVAKAAVNGIDVFRIFDAMNDVRNLRVSIEAVKAAGKHAQGTICYTTSPVHTIE
                     AFVAQGKAMADMGVDSIAIKDMAGLLTPYATGDLVKALKDALPLDVVVHSHDTAGVAS
                     MCQLKAVENGADRIDTAISSMAWGTSHPGTESMVAALRGTPYDTGLDLELIQEIGMYF
                     HAVRKKYHQFESEFTGVDTRVQVNQVPGGMISNLANQLKEQGALNRMGEVLEEIPRVR
                     ADLGFPPLVTPTSQIVGTQAFFNVLAGERYKTITNEVKLYLQGRYGKAPGEVNEQLRR
                     QAIGSEEVIDVRPADLIKPELNKLRSEIGNLAKSEEDVLTYAMFPDIGRKFLEEREAG
                     TLKPEELLPIPTGKGVAAAGAEGTPTEFVIDVHGETYRVDITGVGVKSDNKRHFYLSI
                     DGMPEEVVFEPLNEYVAGSASGRKHASEPGHVSTTMPGNIVDVLVKEGDSVKAGQAVL
                     ITEAMKMETEVQAGIAGTVKAIHVAKGDRVNPGEILIEIAG"
     gene            complement(4525. .5940)
                     /gene="PA5436"
     CDS             complement(4525. .5940)
                     /gene="PA5436"
                     /codon_start=1
                     /transl_table=11
                     /product="probable biotin carboxylase subunit of a
                     transcarboxylase"
                     /protein_id="AAG08821.1"
                     /db_xref="GI:9951765"
                     /translation="MIKKILIANRGEIAVRIVRACAEMGIRSVAVYSEADRHALHVKR
                     ADEAHSIGADPLAGYLNPRALVNLAVESGCDALHPGYGFLSENAELAEICAERGIKFI
                     GPSAQVIRRMGDKTEARRSMIAAGVPCTPGTEGNVADLAEALREAERIGYPVMLKATS
                     GGGGRGIRRCNSREELEQAFPRVISEATKAFGSAEVFLEKCIVNPKHIEAQILADSFG
                     NTVHLFERDCSIQRRNQKLIEIAPSPQLTPEQRAYIGDLAVRAAKAVGYENAGTVEFL
                     LADGEVYFMEMNTRVQVEHTITEEITGIDVVREQIRIASGLELSVKQDDIVHRGYALQ
                     FRINAEDPKNNFLPSFGKITRYYAPGGPGVRTDTAIYTGYTIPPYYDSMCLKLIVWAL
                     TWEEALDRGLRALDDMRVQGVKTTAPYYQEILRNPEFRSGQFNTSFVESHPELTQYSI
```

```
                KRNPSHLAIAIATAIAAHAGL"
     gene       6154. .7089
                /gene="PA5437"
     CDS        6154. .7089
                /gene="PA5437"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG08822.1"
                /db_xref="GI:9951766"
                /translation="MRMTLRQLQVFRAVCESRSYSRAAEEMALTQPAVSLQIRQLEEL
                VGQPLFEYVGKKLYLTDAAEALLKTSTDLFQRLESLDMQLSDLKGSLQGQLRLAVESS
                AKYITPHLFKAFQNLHPEVSLSLTVVNRTQVIKRLADNRDDLVIMSLVPQDMALEFLP
                FLNNPIIAVAPPDHPLCNAAKLALKDLEPFPLLVREPGSGTRKACEEFFQQRRAHFSQ
                IQEIASLESLREGVIAGLGLALLPRHAAHQELANGLLRELPVEELPLYRSWCLVHAKG
                KRLSPVAQAFVAFVREERALISQLAGRFAGPGGSS"
     gene       complement(7365. .8246)
                /gene="PA5438"
     CDS        complement(7365. .8246)
                /gene="PA5438"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG08823.1"
                /db_xref="GI:9951767"
                /translation="MNLLQHIAQSRQQLRKSELKVADHVLNDPASVMHSSMAELAHGV
                GVSEPTIVRFCRAIGCSGFQDLKLKLAQSLAAGASFGQFSIHESDSVADFSLKIFDTT
                LHSLMEVREHLDTHALERAIAAIAHAQRVEFYGFGASGAVASDAQHKFFRLLLSAAAY
                SDPHMQAMSAVTLKPSDVAICISQSGRSKDLLITANLVREAGATLITLCPSQTPLADL
                ATVNLAIDVHEDTDIYTPLTSRIAHLVVIDVLAMGVAMARGPDLVNHLKSVKRSLRSL
                RLSPKVMKNQEERGEPS"
     gene       8332. .9798
                /gene="PA5439"
     CDS        8332. .9798
                /gene="PA5439"
                /codon_start=1
                /transl_table=11
                /product="probable glucose-6-phosphate dehydrogenase"
                /protein_id="AAG08824.1"
                /db_xref="GI:9951768"
                /translation="MSTPATPTSCDILVFGGTGDLALHKLLPALYRLHREDRLPADTR
                IFALARSALDNAAFLALAERNVRAVVARSEFAAEQWKGFAARLDYLAMDASQSADFGR
                LARHLKSSEGRVLVHYLATSPSLFAPIAQNLSIAGLAGPQARIVLEKPLGHSLDSARA
                INQSIGRVFDESRVFRIDHYLGKETVQNLMALRFANALFEPVWRAAHIDHVQISVNET
                LGVENRGGYYDHAGAMRDMVQNHLLQLLCLVAMEAPVRFDAEAVRNEKLKVLQALKPI
                SGLDVQDKTVRGQYAAGRIGGQEVPAYYFEKNVDNDSDTETFVALQAEVENWRWAGVP
                FYLRTGKRMARKCSEIVIQFKPVPHSLIDGGGGPANRLWIRLQPEERISVQLMAKTPG
                KGMQLEPVELDLNLAEALSRNKRRWDAYERLLLDVIEGDSTLFMRRDEVEAAWGWVDP
                ILAGWREHYQSPRPYPAGSNGPEQAQLLLELQGRRWLE"
     gene       9879. .11273
                /gene="PA5440"
     CDS        9879. .11273
                /gene="PA5440"
                /codon_start=1
                /transl_table=11
                /product="probable peptidase"
                /protein_id="AAG08825.1"
                /db_xref="GI:9951769"
                /translation="MTATTLRPELLCPAGTLKSMRYAFAYGADAVYAGQPRYSLRVRN
                NDFDHAQLAEGIAEAHALGKRFYVVVNIAPHNAKLKTFLRDLEPVVAMGPDALIMSDP
                GLIMLVRQHFPQMPIHLSVQANAVNWAAVQFWRQQGLTRVILSRELSLEEIEEIRQQV
                PDMELEVFVHGALCMAYSGRCLLSGYINRRDPNQGTCTNACRWEYKAVEGREDELGNV
                VAREPTLGRGAPTEQVFLLEEGNRPGEPMAAFEDEHGTYIMNSKDLRAVQHVERLARI
                GVHSLKIEGRTKSHYYVARTAQVYRQAIDDAAAGKPFDRGLMDTLESLAHRGYTEGFL
                RRHVHDEYQNYQHGHSLSERQQFVGELTGERRDGLAEVEVKNRFQRGDSVELMTPRGN
                LSLRVEDLRNAAGEAIEVAPGNGHVVYLALPEGLDLCYALLMRDLQPGQDTRQPQAER
                SGGCAGKGACGCAH"
BASE COUNT      1912 a    3828 c    3971 g    1949 t
ORIGIN
```

```
Query Match             2.8%;  Score 37.8;  DB 1;  Length 11660;
Best Local Similarity   50.8%;  Pred. No. 1.5e+02;

  Matches   90;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy     338 taagcggcgaagaagtgcgccgcaacctgatccaatccggtttgtggactgcgctgcgca 397
           | | ||||   || |||  |||| | | |||||    || ||   ||   | ||||
Db    5039 TCACCGGCATCGACGTGGTCCGCGAGCAGATCCGCATCGCTTCCGGGCTGGAGCTGTCGG 4980

Qy     398 cccgtccgttcagcaaaattcctgccgtcgatgccgagccgttcgccatcttcgtcaatg 457
           |   | | |  ||   | | |  || | | |||  || ||||  ||||  || | | |
Db    4979 TCAAGCAGGACGACATCGTCCATCGCGGCTACGCCCTGCAGTTCCGCATCAACGCCGAGG 4920

Qy     458 cgatggacaccaatccgctggctgccgaccctacggtcattatcaaagaagccgccg 514
             | | | |||    ||| |  |   |   | | |||    | |  | ||| |||
Db    4919 ACCCGAAGAACAACTTCCTGCCCTCGTTCGGCAAGATCACCCGCTACTACGCCCCCG 4863


RESULT  35
AC005499/c
LOCUS       AC005499              103573 bp    DNA     linear   PLN 05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 208 of 255 of the
            complete sequence. Sequence from clones T19C21, T6A23.
ACCESSION   AC005499 AE002093
VERSION     AC005499.2  GI:6598470
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 103573)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20083487
   PUBMED   10617197
REFERENCE   2  (bases 1 to 103573)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:3785992.
            The sequence and annotation of chromosome 2 were merged from those
            of the individual clones on this chromosome after removing
            overlaps. For detailed information, please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).

            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
            complete sequence against a peptide database and plant EST
            databases at TIGR, and manual curations based on those analyses.
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by two
            or more gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
            numbered from the top to bottom of the chromosome.
```

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES Location/Qualifiers

source 1. .103573
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"

misc_feature complement(<1. .2468)
/note="Sequence from clone T19C21"

mRNA join(<1367. .1588,1838. .2044,2114. .2646,2737. .2898, 2995. .>3409)
/gene="At2g38470"

gene <1367. .>3409
/gene="At2g38470"
/note="T19C21.4"

CDS join(1367. .1588,1838. .2044,2114. .2646,2737. .2898, 2995. .3409)
/gene="At2g38470"
/codon_start=1
/product="putative WRKY-type DNA-binding protein"
/protein_id="AAC67339.2"
/db_xref="GI:6598471"
/translation="MDNSRTRQNMNGSANWSQQSGRTSTSSLEDLEIPKFRSFAPSSI SISPSLVSPSTCFSPSLFLDSPAFVSSSANVLASPTTGALITNVTNQKGINEGDKSNN NNFNLFDFSFHTQSSGVSAPTTTTTTTTTTTTNSSIFQSQEQQKKNQSEQWSQTETR PNNQAVSYNGREQRKGEDGYNWRKYGQKQVKGSENPRSYYKCTFPNCPTKKKVERSLE GQITEIVYKGSHNHPKPQSTRRSSSSSSTFHSAVYNASLDHNRQASSDQPNSNNSFHQ SDSFGMQQEDNTTSDSVGDDEFEQGSSIVSRDEEDCGSEPEAKRWKGDNETNGGNGGG SKTVREPRIVVQTTSDIDILDDGYRWRKYGQKVVKGNPNPRSYYKCTTIGCPVRKHVE RASHDMRAVITTYEGKHNHDVPAARGSGYATNRAPQDSSSVPIRPAAIAGHSNYTTSS QAPYTLQMLHNNNTNTGPFGYAMNNNNNNSNLQTQQNFVGGGFSRAKEEPNEETSFFD SFMP"

repeat_region complement(1497. .1622)
/rpt_family="(GAAAA)n"

repeat_region 1980. .2021
/rpt_family="(CAA)n"

misc_feature complement(2469. .>103573)
/note="Sequence from clone T6A23"

mRNA complement(join(<3830. .4021,4103. .4226,4314. .4637, 5922. .>6013))
/gene="At2g38480"

gene complement(<3830. .>6013)
/gene="At2g38480"
/note="T6A23.32; predicted by genscan"

CDS complement(join(3830. .4021,4103. .4226,4314. .4637, 5922. .6013))
/gene="At2g38480"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAC67370.1"
/db_xref="GI:3786024"
/translation="MQIKLVDVIEKEREDCYKAISAVGNYVLSVCLLFLVPNLRIKNR TNSDQSFEICNMTNPDNMKPVEATDVESAAEKTSEPTPASGTSTITQRWKREDLIKKA SPITRGICLLFSLIAFLIMVSNKHGYGRNFNDYEEYRYVLAISIISTLYTAWQTFAHF SKREIFDRRTSILVDFSGDQIVAYLLISAASSAIPLTNIFREGQDNIFTDSAASAISM AIFAFIALALSALFSGYKLSTHSFI"

mRNA complement(<6779. .>8074)
/gene="At2g38490"

gene complement(<6779. .>8074)
/gene="At2g38490"
/note="T6A23.31; contains a protein kinase domain profile (PDOC00100)"

CDS complement(6779. .8074)
/gene="At2g38490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC67369.1"
/db_xref="GI:3786023"
/translation="MAEDSNSSESIIVNVTGDDNKSALFGKYDLGKLLGSGAFAKVYQ AEDLQNGGESVAIKVVQKKRLKDGLTAHVKREISVMRRLRHPHIVLLSEVLATKTKIY FVMELAKGGELFSRVTSNRFTESLSRKYFRQLISAVRYCHARGVFHRDLKPENLLLDE NRDLKVSDFGLSAMKEQIHPDGMLHTLCGTPAYVAPELLLKKGYDGSKADIWSCGVVL FLLNAGYLPFRDPNIMGLYRKIHKAQYKLPDWTSSDLRKLLRRLLEPNPELRITVEEI LKDPWFNHGVDPSEIIGIQADDYDLEENGKILNAFDLISSASSSNLSGLFGNFVTPDH CDQFVSDESTAVIMRKVEEVAKQLNLRIAKKKERAIKLEGPHGVANVVVKVRRLTNEL VMVEMKNKQRDVGLVWADALRQKLRRLINQPVYKVPDKP"

repeat_region 8079. .8124
/rpt_family="(CAAAA)n"

mRNA complement(join(<10806. .11136,11403. .>11911))
/gene="At2g38500"

gene complement(<10806. .>11911)
/gene="At2g38500"
/note="T6A23.30; predicted by genefinder"

CDS complement(join(10806. .11136,11403. .11911))
/gene="At2g38500"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAC67368.1"
/db_xref="GI:3786022"
/translation="MALMRTRSQLNVSSLTPPPPPPSPIPRARGSRCAASEILTEIIE RSIQVPELTLPESHSGGESCGSRHLIPAEIDFRLLASRREGSVDRLVRSAREFGAFRV SYHGISGEELRSLVRESGRVFGVLEGRDTGFHRSVVGNRDEIVWVRSWKERMEWAREY IGPERYRCFSQEMENVADKLEDIARKLGQIMVENSRRPNDKKIQRGESVLSVYRYNHE NVTEQSPPLPKERTEEMLHYTLSLHLPAKNCEFRVNSGKGPLSFHADPDTILVTFGRQ LEV"

repeat_region 14810. .14882
/rpt_family="POLY_A"

repeat_region complement(15232. .15278)
/rpt_family="(GAAAA)n"

repeat_region complement(15653. .15723)
/rpt_family="(TA)n"

mRNA <16855. .>18315
/gene="At2g38510"

gene <16855. .>18315
/gene="At2g38510"
/note="T6A23.29; predicted by grail"

CDS 16855. .18315
/gene="At2g38510"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAC67367.1"
/db_xref="GI:3786021"
/translation="MQVGEEMASLTKIACPIVMTSLLIFSRSIISMWFLSHLGKVELA GGALAMGFGNITGVSVLKGLSVGMDPICGQAFGAKRWTVLSHTFQKMFCLLIVVSVPI AVTWLNIEPIFLRLGQDPDITKVAKTYMLFFVPELLAQAMLHPLRTFLRTQGLTSPLT ISAIVSILLHPLFNYVFVVRMRLGVKGVAIAMAFNTMNIDVGLLVYTCFSDSLIKPWE GLALRSLFRGWWPLLSLAAPSAISVCLEYWWYEIMLFLCGLLGNPKASVAAMGILIQT TGILYVVPFAISSAIATRVGHALGGGQPTRAQCTTVIGLILAVAYGLAAAVFVTALRS VWGKMFTDEPEILGLISAALPILGLCEIGNSPQTAACGVLTGTARPKDGARVNLCAFY IVGLPVAVTTTFGFKVGFRGLWFGLLSAQMTCLVMMLYTLIRTDWSHQVKRAEELTSA AADKSHSEDETVHAEVQDDDDVSSNDLEIGLLQNTN"

mRNA <18795. .>19448
/gene="At2g38520"

gene <18795. .>19448
/gene="At2g38520"
/note="T6A23.28"

CDS 18795. .19448
/gene="At2g38520"
/codon_start=1
/product="putative retroelement pol polyprotein"
/protein_id="AAC67366.1"

Query Match 2.8%; Score 37.8; DB 8; Length 103573;

```
  Best Local Similarity   55.8%;  Pred. No. 1.7e+02;
  Matches   72;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy    128 tgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagacaaaaaga 187
          | ||||  |   |   |  ||  || ||||||    ||   ||||    |   |  |||
Db  76702 TCAAAGCTACACAGATCTCTGAAATCGAAAAAGCTTGAGCTGTGTTAATGGCGGAGGAGA 76643

Qy    188 atccgggcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcaccgtggcg 247
          |   ||  ||| || | || |||| | || ||   || |||||| |||| |   | | |
Db  76642 AATGGGTGGTGATGGTGACGGCGCAGACTCCAACGAATATCGCCGTGATTAAGTATTGGG 76583

Qy    248 aaaagcgcg 256
            |||| | |
Db  76582 GAAAGAGAG 76574


RESULT  36
AE002563/c
LOCUS       AE002563               10944 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 205 of 206
            of the complete genome.
ACCESSION   AE002563 AE002098
VERSION     AE002563.1  GI:7227405
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10944)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10944)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10944
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            268. .1233
                     /gene="NMB2148"
     CDS             268. .1233
                     /gene="NMB2148"
                     /note="similar to SP:P37247 PID:1197006 percent identity:
                     61.94; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase, IS30 family"
                     /protein_id="AAF42456.1"
                     /db_xref="GI:7227406"
                     /translation="MSYTQLTQGERYHIQYLSRHCTVTEIAKQLNRHKSTISREIRRH
                     RTQGQQYSAEKAQRQSQTIKQRKRQPYKLDSQLIQHIDTLIRRKLSPEQVCAYLCKHH
                     QITLHHSTIYRYLRQDKSNGSTLWQHLRICSKPYRKRYGSTWTRGKVPNRVGIENRPA
                     IVDQKSRIGDWEADTIVGKGQKSALLTLVERVTRYTIICKLDSLKAEDTARAAVRALK
                     AHKDRVHTITMDNGKEFYQHTKITKALKAETYFCRPYHSWEKGLNENTNGLIRQYFPK
                     QTDFRNISDREIRRVQDELNHRPRKTLGYETPSVLFLNLFQPLIH"
     gene            complement(1322. .1582)
                     /gene="NMB2149"
     CDS             complement(1322. .1582)
                     /gene="NMB2149"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF42457.1"
                     /db_xref="GI:7227407"
                     /translation="MRQISLTDYFCKGLYLFDIFEDIVFLGKRLFTVRKFLDFSELLE
                     MVNTQGYFLEFLYLYSGLNLNQYGVASPCRTICTVCGFVALS"
     gene            complement(1967. .2536)
                     /gene="NMB2150"
     CDS             complement(1967. .2536)
                     /gene="NMB2150"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF42458.1"
                     /db_xref="GI:7227408"
                     /translation="MLFPRIIAASAQRKLSVYLKKGGLVAYPTESCYGLGCLPTLAKA
                     LGKLAHLKKRPQHKGMIVIGNQFEQLQPLLQMPSENLQDMLRKEWPAPKTFLLSAKSC
                     VLPELRGKQRSKLAVRVPAHVGARRLCQALQTPLVSTSCNRAGKRACRTEREVRRQFG
                     RDVWIVGGRIGRQKSPSQIIDGETGKRLR"
     gene            complement(2603. .3874)
                     /gene="NMB2151"
     CDS             complement(2603. .3874)
                     /gene="NMB2151"
                     /note="similar to GB:M66160 SP:P26977 PID:154288 percent
                     identity: 75.06; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoribosylamine--glycine ligase"
                     /protein_id="AAF42459.1"
                     /db_xref="GI:7227409"
                     /translation="MKLLVIGNGGREHALAWKLAQSPKVETVFVAPGNAGTAIEPKLQ
                     NIDLTAHQDLIEFCRKENIVFTVVGPEAPLAAGIVDDFRAAGLKIFGPTQYAAQLESS
                     KDFAKAFMAKYNIPTAQYQTFENADAAHDYVNQKGAPIVIKADGLAAGKGVIVAMTLD
                     EAHAAIDDMLLDNKMGNAGARVVIEDFLQGEEASFIVMVDGNNVLPMATSQDHKRLLD
                     GDKGLNTGGMGAYSPAPVVTPVVYERAMNEIILPTVAGMKAEGHEFTGFLYAGLMIDQ
                     SGAPYTIEFNCRFGDPETQPIMSRLNSDLSDLVEAAIDGKLDSVTAEWSPQTAVGVVL
                     AAQNYPETPKKGDIISGLDAANQIGKVFHAGTTANEKGDVLTNGGRVLCVVGLGDNVA
                     QAKAKAYGALEKISFDGMQYRKDIADKAINR"
     gene            4032. .4349
                     /gene="NMB2152"
     CDS             4032. .4349
                     /gene="NMB2152"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF42460.1"
                     /db_xref="GI:7227410"
                     /translation="MQKRIDEIQSKYREWCHLLPQLEEDIRRWKHVVTLIRDMDNFYT
                     HEYQACHQAIEDGVELDLSTEGEYSIMSEDALWNALGEFHQLAWLYLRSSVDALDKYT
                     QED"
     gene            complement(4401. .5012)
                     /gene="NMB2153"
     CDS             complement(4401. .5012)
                     /gene="NMB2153"
                     /note="conserved hypothetical protein; identified by
```

```
                Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF42461.1"
                /db_xref="GI:7227411"
                /translation="MITTYKTITSLTQAEFKDKGSRFIAFAYPIRTLADVKKYLDPLK
                EEHHKARHWCYAYRLGVDGVQFRANDDGEPSGSAGRPILGQIDSVGITDVLVVVVRYF
                GGTLLGVPGLIHAYKEATAQALAVAEVVEKNIEKTVWLKCEYPFLNEAIRIAKQYQAD
                ILERDLQLDCRLTVSLSLVNYEACVAAWKNTRQIEVNTEKPFE"
gene            complement(5075. .6010)
                /gene="NMB2154"
CDS             complement(5075. .6010)
                /gene="NMB2154"
                /note="similar to SP:P53573 PID:1209055 percent identity:
                74.19; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="electron transfer flavoprotein, alpha subunit"
                /protein_id="AAF42462.1"
                /db_xref="GI:7227412"
                /translation="MSVLIIVEHDNKQLNPTTLHAVTAAAKLGKVDLLVAGNGASAVV
                EFAKQVAGVKKVLVADAAHYAEGLAEELAPLVVKLAADYRYVAATATTFGKNLLPRVA
                ALLDVPQISDLTEIVDNTTFVRPIYAGNAFETVQADSEKLVLTFRATVFDAVAAQGGN
                AEVINVEATPAQNLSRFVNRQLSHSDRPELTQAKVIVSGGRALGSAEKFNEVLTPLAD
                VLGAAIGASRAAVDAEYAPNDAQVGQTGKVVAPQLYFAIGISGAIQHVAGMQDSKVIV
                AINKDADAPIFNVADYGLVGDLFEIVPQLTEVLKN"
gene            complement(6021. .6770)
                /gene="NMB2155"
CDS             complement(6021. .6770)
                /gene="NMB2155"
                /note="similar to PID:1173889 percent identity: 98.80;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="electron transfer flavoprotein, beta subunit"
                /protein_id="AAF42463.1"
                /db_xref="GI:7227413"
                /translation="MKALVAVKRVVDYNVKVRVKADGSDVDIGNVKMSMNPFDEIAVE
                EAVRLKEAGKVSEIVAVSLGEKKCEETLRTALAMGADRAIHVETDTKLESLAVAKLLK
                AVADKENPQIFFLGKQAIDDDANQVAQMLAALLNAAQGTFASKVQIEGDEVQIVREID
                GGEETIALKLPAVISADLRLNEPRFVKLPNIMAAKKKPLEKLTPDDLVADISPRLKTV
                KFAEPKARQAGVKVASVAELVEKLKNEAKVI"
gene            complement(6997. .7965)
                /gene="NMB2156"
CDS             complement(6997. .7965)
                /gene="NMB2156"
                /note="similar to PID:1016752 percent identity: 97.83;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="heptosyltransferase I"
                /protein_id="AAF42464.1"
                /db_xref="GI:7227414"
                /translation="MKILLVRLSSMGDLIHTLPAIEDLARQCPDVELHWLCEAGFADI
                ARLHPFVKKIHVMKWRQWRKHLFRAETWREMGHLKQTLRQEVFDFVLDSQGLIKSACF
                AKMAKSPICGLDKNSAREGWAALAYVKTYAVPKGKNAVWRNRELFAQVFGYVMPETQV
                FGLTVPEAGRLKNLAQPYYAALHATSRDSKLWPVENWRSLLQKLNEEQQCNIYLPWGN
                EDEKTRAKQIADGLPFAIVCDKINLLQAAYLLKYAVGIVGVDTGLLHLANALEKPVVG
                IYTDTDPIKTGVQVSPIAKNLGNIGQIPTADLVYQALMDCVAADKG"
gene            complement(8002. .8637)
                /gene="NMB2157"
CDS             complement(8002. .8637)
                /gene="NMB2157"
                /note="similar to PID:1399892 GB:AL123456 percent
                identity: 49.70; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="pyrazinamidase/nicotinamidase PncA, putative"
                /protein_id="AAF42465.1"
                /db_xref="GI:7227415"
```

```
  Query Match             2.8%;  Score 37.6;  DB 1;  Length 10944;
  Best Local Similarity   52.6%;  Pred. No. 1.7e+02;
  Matches   82;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     870 cgaattggttgacacagacaaccgcgtgatttccggttcggtattgaacggcgcgattac 929
           ||||||||||  ||    | | |||    |  ||       | | | |||| ||||||
Db    9965 CGAATTGGTTTCCATCCTCGAACGCCCTGTAACCAAAGAAGAAATCAACGCCGCGATGAA 9906

Qy     930 acaaggcgcgcacgattatttgggacgctaccacaatcagatttccgttatcgaagaagg 989
           |   | |||   ||| | ||  ||   | || |  |||| ||       |  ||| |  |
Db    9905 AGCCGCCGCCAGCGAATCTTACGGCTACAACGAAGATCAAATCGTTTCTTCCGACGTTGT 9846

Qy     990 ccgcagcaaagagctgttcggctgggttgcgccgca 1025
           | ||| | || |    |    ||  | |||| | ||
Db    9845 CGGCATCGAATACGGCTCGCTCTTCGATGCGACCCA 9810


RESULT  37
AX044035/c
LOCUS       AX044035              172325 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 114 from Patent WO0066791.
ACCESSION   AX044035
VERSION     AX044035.1  GI:11342919
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 172325)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 114 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .172325
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    43072 a  47583 c  41465 g  40205 t
ORIGIN


  Query Match             2.8%;  Score 37.6;  DB 6;  Length 172325;
  Best Local Similarity   52.6%;  Pred. No. 2e+02;
  Matches   82;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     870 cgaattggttgacacagacaaccgcgtgatttccggttcggtattgaacggcgcgattac 929
           ||||||||||  ||    | | |||    |  ||       | | | |||| ||||||
Db  168435 CGAATTGGTTTCCATCCTCGAACGCCCTGTAACCAAAGAAGAAATCAACGCCGCGATGAA 168376

Qy     930 acaaggcgcgcacgattatttgggacgctaccacaatcagatttccgttatcgaagaagg 989
           |   | |||   ||| | ||  ||   | || |  |||| ||       |  ||| |  |
Db  168375 AGCCGCCGCCAGCGAATCTTACGGCTACAACGAAGATCAAATCGTTTCTTCCGACGTTGT 168316

Qy     990 ccgcagcaaagagctgttcggctgggttgcgccgca 1025
           | ||| | || |    |    ||  | |||| | ||
Db  168315 CGGCATCGAATACGGCTCGCTCTTCGATGCGACCCA 168280


RESULT  38
AL646061
LOCUS       AL646061              198050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 chromosome, complete sequence;
            segment 5/19.
ACCESSION   AL646061 AL646052
```

```
VERSION     AL646061.1  GI:17427781
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 198050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .198050
                     /organism="Ralstonia solanacearum"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     gene            151. .642
                     /gene="RSc0770"
                     /note="RS05084"
     CDS             151. .642
                     /gene="RSc0770"
                     /EC_number="1.-.-.-"
                     /function="small molecule metabolism"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (SMALL
                     SUBUNIT) OXIDOREDUCTASE PROTEIN"
                     /protein_id="CAD14472.1"
                     /db_xref="GI:17427782"
                     /translation="MHTRPPSSDGDDPAGLTRRQWLQSALAATAASLVGSVTLYAVAQ
                     APGEPLDAFMGLSQALTARSALDRAVGARLLAALQQASAGFAAQLRPLAQSLAAGTLS
                     GLQQETALRILQAWYTGVVDGTVVTYEQALMYGVVSDILVIRTYCPNRPGFWAEPPVE
                     KQS"
     gene            642. .2261
                     /gene="RSc0771"
                     /note="RS05083"
     CDS             642. .2261
                     /gene="RSc0771"
                     /EC_number="1.-.-.-"
                     /function="small molecule metabolism"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (LARGE
                     SUBUNIT) OXIDOREDUCTASE PROTEIN"
                     /protein_id="CAD14473.1"
                     /db_xref="GI:17427783"
                     /translation="MAETRQAEQADIVVVGSGVAGALVAYELARAGKSVLMLEAGPRL
                     PRWEIVERFRNQADKMDFMAPYPSTPWAPHPEYGPDNNYLILKGEHKFDSQYIRAVGG
                     TTWHWAASTWRFLPNDFKLRSVYGIARDWPIQYDDLERYYGRAEAALGVWGPNDEELG
                     SPRSQPYPMAPLPLSFNERTVKDALNAHDPALRVVTEPVARNSRPYDGRPTCCGNNNC
                     MPICPIGAMYNGIVHVEKAEQAGARLIENAVVYKLEVGAGKRIVAAHYKDPKGVDHRV
                     EGKWFVLAANGIETPKLMLMSTGADFPKGVGNRSDMVGRNLMDHPGTGVSFYADRKLW
                     PGRGPQEMTSLIGFRDGPFRATQAGKKLHLSNISRIEQETTRIFKAGKLLKPAELDAR
                     IRDQAARYVQFDSFHEILPLPENRIVPSATGTDALGIPRPEITYRIDDYVKRSAVHTR
                     EVYATAARVLGGTDVQFHDDFAPNNHITGATIMGSDPKDSVVDKDCRTFDHPNLFISS
                     SATMPTVGTVNVTLTIAALALRIADQLKKEA"
gene                 2263. .3552
                     /gene="RSc0772"
                     /note="RS05082"
CDS                  2263. .3552
                     /gene="RSc0772"
                     /EC_number="1.-.-.-"
                     /function="small molecule metabolism; energy transfer;
                     electron transport"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PUTATIVE OXIDOREDUCTASE DEHYDROGENASE
                     (CYTOCHROME C SUBUNIT) SIGNAL PEPTIDE PROTEIN"
                     /protein_id="CAD14474.1"
                     /db_xref="GI:17427784"
                     /translation="MRRAKAASFLLLSVLGAAAAARDADAPADAAQIKRGEYLAIAGD
                     CVACHTAPGGKPFAGGLPLQIPMLGTIYSSNITPDPKTGIGKWRYEDFERAVRQGVSD
                     DGENLYPAMPYPSYAKINDADMRDLYAYFRYGVAAVQYAPPPGTIRWPLNMRWPLKLW
                     NLAFLKRAPYLPKGDMTPTWNRGAYLAQGLAHCGTCHTPRGFAMQEKAMDERGKGYLA
                     GSTLAGWSAYNITSDPVSGIGGWKPEQIVQYLRTGSVPGIGQAAGPMGEAVQHSFSRM
                     TERDVRAIAEYLRTVPAVSTGAQRARHGWGTPATDVIALRGKPIETTIDAARLYLGNC
                     ATCHQADGRGTPDGYYPPLLHNSAVGARDTRNLVQVMLHGIERRAGDRHIGMPAFGRQ
                     LSDAQLAALANYVTRQFGDPATPALTADEIAKRRMPQ"
gene                 complement(3559. .5781)
                     /gene="RSc0773"
                     /note="RS05081"
CDS                  complement(3559. .5781)
                     /gene="RSc0773"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD14475.1"
                     /db_xref="GI:17427785"
                     /translation="MRTSARVILGLGITPIVIVAAAIVFVLTFDWNRAKPYLNDHVSQ
                     AVGRPFAINGDLALTWKKPEGESGWRAYVPWPRLIASDIAIGNPDWARQPHVATVQRL
                     TFVLEALPLLAHRIVIPSVTLDTPMIALERDSQNRNNWTFHFGAQDSQPSDWKLELRE
                     IAFSKGSVQLDDEVKRIHLAATLDTVDNQALYTAASGTAIRGSEPAAPSASGAAASTA
                     PAAPDPQQPYGIAWTVKGTYNNAAIRGSGKAGGVLRLQDARRPYPLQADVTVGKTRID
                     LAGTLTNPSSLTALDLRLHLSGASMAHLYPLTGVVLPDTPPFDTRGRLIGELRRQGSS
                     WRYEQFTGRVGGSDLGGTLTFAMRPEPGQRPQLTGELVSHQLLFADLAPIIGADSNAS
                     KQRRDAPVRQPADKVLPVEPFRTDRWNAIDADVKFTGERIVRTADLPIDHLVTHIKLQ
                     DAVLTLDPLNFGVAGGTLTSNLRLDGQSAPMQASAALAARHFKIRQLFPNIESMHASI
                     GEINGDARLSATGNSVAALLGASNGELKILVEQGTVSKFILEAMGLNVGNVILTKLFG
                     DKQVSINCAAGDFAVSDGLAQARTFVVDTQDAVIDVTGATSFKHESLDFTIHPDSKGL
                     RVFSLRTPLYVKGTYKHPDVSVNPAVVALRAGGAVALAFAAPVAAVLPVLELQPAPDS
                     PCGKLLADVRQRPTAPPPGKVYRSGQTGRAADGSPKDPAAAPARQPALPLRDPTTAGG
                     "
gene                 5883. .6785
```

```
                     /gene="RSc0774"
                     /note="RS05080"
     CDS             5883. .6785
                     /gene="RSc0774"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD14476.1"
                     /db_xref="GI:17427786"
                     /translation="MSHVAISAPAHPPALDDEYARRFGGVARLYGPQALERFAAAHIC
                     VIGVGGVGSWVAEALARCGIGRLTLIDLDHIAVSNTNRQIHALGDAYGKAKVDAMAER
                     IAEINPRAEISRIDDFVTVENVDALLGHPFDYVVDAIDAVKVKTAIVAFCQRTGTRVV
                     TCGAAGGQLDPTRIRVADLARTIQDPLLAKVRGNLRRQHGFPRDPKARFGIDAVYSDE
                     PLRYPEPEQQACAVEVPAESSHAGIDDLAAAGAIEVVNAPPAPQPSRGPQGLACAGFG
                     SSVCVTAVFGMVAASTALRAIAAV"
     gene            complement(6890. .8257)
                     /gene="katGb"
                     /note="RSc0775; RS05079"
     CDS             complement(6890. .8257)
                     /gene="katGb"
                     /EC_number="1.11.1.6"
                     /function="cell processes; protection responses;
                     detoxification"
                     /note="Product confidence : probable
                     Gene name confidence : putative
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE PEROXIDASE/CATALASE (CATA)
                     OXIDOREDUCTASE PROTEIN"
                     /protein_id="CAD14477.1"
                     /db_xref="GI:17427787"
                     /translation="MGPEPEAAGIEEQGLGWSSRFGTGKGTDAITSGLEVTWTTTPTQ
                     WSNNFFQNLFGYEWELTKSPAGAHQWVAKDAAETIPDAHDPSAKHLPTMLTTDLSLRF
                     DPAYEKISRRFYEHPEQFADAFARAWFKLTHRDMGPRARYLGPEVPAEALIWQDPIPA
                     VDHKLIDTQDIAALKARILASGLSVSQLVSTAWASASTFRGSDMRGGANGARIRLAPQ
                     KDWEANQPAQLAKVLETLESIQGAFNGAQSGGKKVSLADLIVLAGCAGVEQAAKNAGH
                     AVEVPFTPGRMDAAQAQTDVESFAVLEPIADGFRNYQKGKYTLPAEALLVDKAQLLTL
                     TAPEMTVLVGGLRVLDTNIGQTRHGVFTQRPESLTNDFFVNLLDMGTEWKATDGRDVF
                     EGRDRATGALKWTGTRVDLVFGSHSQLRALAEVYGSADAQAKFVRDFVAAWDKVMNLD

  Query Match              2.8%;  Score 37.6;  DB 1;  Length 198050;
  Best Local Similarity   51.2%;  Pred. No. 2e+02;
  Matches   88;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy   1150 gtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtcggcgatacc 1209
          ||| ||  | || |  ||| ||   ||  ||  ||  || |     |||||| ||| |||
Db 121340 GTGCTGGGCGTGCAGTTCCGGCTATCCAAGCGGTTCGGCAAGAACTTCGTCGACGACACC 121399

Qy   1210 gacagcgcgcaggcattgggttgcttggaattggacgaagaagacctcgctttgtgcagc 1269
          || |  | |||     | |   |       |||  ||||  | |||  |    ||
Db 121400 GAGATTCCCCAGCAGCGGCGCAACAACACGGTGGCGGAAGCCGGCCTGACGGACCGCCAT 121459

Qy   1270 ttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctgg 1321
          | | || |  | |   | | |||   | ||| |||| | ||| | |   |||
Db 121460 TACTTCGGGGCCGCGCAGTTCGACGGCAGCCTGCTGCTCCGCCACGGCGTGG 121511


RESULT 39
NMA1Z2491/c
LOCUS       NMA1Z2491              340806 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 1/7.
ACCESSION   AL162752 AL157959
VERSION     AL162752.2  GI:7378778
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 340806)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 340806)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .340806
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            complement(38. .358)
                     /gene="NMA0001"
     CDS             complement(38. .358)
                     /gene="NMA0001"
                     /note="NMA0001, unknown, len: 106 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0001"
                     /protein_id="CAB83321.1"
                     /db_xref="GI:7378779"
                     /db_xref="SPTREMBL:Q9JX93"
                     /translation="MSNTQTIRSADFTTSRAWGALDIANMNGTTVRLHWTDQPYKWHV
                     NDGEEVFAVMDGEVDMHYRENGEEHIVRLKSGDIFYAGIGTEHVAHPRGETRILVIEK
                     EGSV"
     gene            complement(414. .2438)
                     /gene="nuoL"
     CDS             complement(414. .2438)
                     /gene="nuoL"
                     /EC_number="1.6.5.3"
                     /note="NMA0002, nuoL, NADH dehydrogenase I chain L, len:
                     674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase
                     I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to
                     entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone
                     (complex I), various chains, and to entry PF00662
                     oxidored_q1_N, NADH-Ubiquinone oxidoreductase (complex I),
                     chain 5 N-terminus"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain L"
                     /protein_id="CAB83322.1"
                     /db_xref="GI:7378780"
                     /db_xref="SWISS-PROT:Q9JX92"
                     /translation="MNDMTLYLIIALVPLAGSLIAGLFGNKIGRAGAHTVTILGVAVS
                     AVLSAYVLWGFLNGSRAKFDENVYTWLTMGGLDFSVGFLVDTMTAMMMVVVTGVSLMV
                     HIYTIGYMHDEKVGYQRFFSYISLFTFSMLMLIMSNNFIQLFFGWEAVGLVSYLLIGF
                     YFKRPSATFANLKAFLINRVGDFGFLLGIGLVLAYFGGSLRYQDVFAYLPNVQNATIQ
                     LFPGVEWSLITVTCLLLFVGAMGKSAQFPLHVWLPDSMEGPTPISALIHAATMVTAGL
                     FMVSRMSPIYEMSSTALSVIMVIGAITALFMGFLGVIQNDIKRVVAYSTLSQLGYMTV
                     ALGASAYSVAMFHVMTHAFFKALLFLAAGSAIIGMHHDQDMRHMGNLKKYMPITWLTM
```

```
                LIGNLSLIGTPFFSGFYSKDSIIEAAKYSTLPGSGFAYFAVLASVFVTAFYAFRQYFM
                VFHGEEKWRSLPEHHSDGHGEEHHGLGKNDNPHESPLVVTLPLILLAVPSVIIGYIAI
                EPMLYGDFFKDVIFVNADAHPTMHIMKEEFHGALAMVSHSLHSPVLYLAIAGVLSAWL
                LYVKLPHLPAKIAQAFRPVYVLFENKYYLDALYFNVFAKGTRALGTFFWKVGDTAIID
                NGIVNGSARLVGAVAAQVRKVQTGFIYTYAAAMVFGVLVLLGMTFWGLFR"
misc_feature    1035. .1044
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    complement(1116. .2030)
                /gene="nuoL"
                /note="Pfam match to entry PF00361 oxidored_q1,
                NADH-Ubiquinone/plastoquinone (complex I), various chains,
                score 351.80, E-value 7.3e-102"
misc_feature    complement(2061. .2243)
                /gene="nuoL"
                /note="Pfam match to entry PF00662 oxidored_q1_N,
                NADH-Ubiquinone oxidoreductase (complex I), chain 5
                N-terminus, score 77.20, E-value 3.5e-19"
gene            complement(2487. .2828)
                /gene="NMA0003"
CDS             complement(2487. .2828)
                /gene="NMA0003"
                /note="NMA0003, unknown, len: 113 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0003"
                /protein_id="CAB83323.1"
                /db_xref="GI:7378781"
                /db_xref="SPTREMBL:Q9JX91"
                /translation="MMDKNQLEQEFHKAMLNIYQEALNLPQPYKATRFLQIVNEFGGK
                EAADKLLSTGEKKTQTGFTELILSGGGVHALKYSMEYLVLQKPWCDLFTEEQLAVARK
                RLERVGFVSPK"
RBS             complement(2833. .2837)
gene            complement(2952. .3527)
                /gene="NMA0004"
CDS             complement(2952. .3527)
                /gene="NMA0004"
                /note="NMA0004, unknown, len: 191 aa; similar to
                hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa),
                fasta scores; E(): 0; 80.5% identity in 190 aa overlap,
                and to proteins involved in cell cycle e.g. TR:O54679
                (EMBL:AF036487) Lactococcus lactis Plasmid pNZ4000
                putative mobilization protein (200 aa), fasta scores; E():
                0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell
                filamentation protein FIC (200 aa), fasta scores; E():
                0.093, 29.4% identity in 126 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0004"
                /protein_id="CAB83324.1"
                /db_xref="GI:7378782"
                /db_xref="SPTREMBL:Q9JQR9"
                /translation="MPSENPIGKTMKSIDEQSLHNARRLFESGDIDRIEVGTTAGLQQ
                IHRYLFGGLYDFAGQIREDNISKGGFRFANAMYLKEALVKIEQMPERTFEEIIAKYVE
                MNIAHPFLEGNGRSTRIWLDLVLKKNLKKVVNWQNVSKTLYLQAMERSPVNDLELRFL
                LKDNLTDDVDNREIIFKGIEQSYYYEGYEKG"
misc_feature    complement(3516. .3525)
                /gene="NMA0004"
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(3561. .3866)
                /gene="nuoK"
CDS             complement(3561. .3866)
                /gene="nuoK"
                /EC_number="1.6.5.3"
                /note="NMA0005, nuoK, NADH dehydrogenase I chain K, len:
                101 aa; simlar to many e.g. NUOK_RHOCA NADH dehydrogenase
                I chain K (EC 1.6.5.3) (102 aa), fasta scores; E():
                1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam
                match to entry PF00420 oxidored_q2,
                NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
                /codon_start=1
                /transl_table=11
                /product="NADH dehydrogenase I chain K"
                /protein_id="CAB83325.1"
                /db_xref="GI:7378783"
                /db_xref="SPTREMBL:Q9JQU9"
                /translation="MITLTHYLVLGALLFGISAMGIFMNRKNVLVLLMSIELMLLAVN
                FNFIAFSQHLGDTAGQIFVFFVLTVAAAESAIGLAIMVLVYRNRQTINVADLDELKG"
misc_feature    complement(3564. .3863)
                /gene="nuoK"
                /note="Pfam match to entry PF00420 oxidored_q2,
                NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
                score 141.50, E-value 1.5e-38"
gene            complement(3863. .4534)
                /gene="nuoJ"
CDS             complement(3863. .4534)
                /gene="nuoJ"
                /EC_number="1.6.5.3"
                /note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
                223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
                I chain J (EC 1.6.5.3) (202 aa), fatsa scores; E():
                9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
                match to entry PF00499 oxidored_q3,
                NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
                /codon_start=1
                /transl_table=11
                /product="NADH dehydrogenase I chain J"
                /protein_id="CAB83326.1"
                /db_xref="GI:7378784"
                /db_xref="SPTREMBL:Q9JX90"
                /translation="MTFQLILFYIFAVIILYGAIKTVTAKNPVHAALHLVLTFCVSAM
                LWMLMQAEFLGVTLVVVYVGAVMVLFLFVVMMLNIDIEEMRAGFWRHAPVAGVVGTLL
                AVALILILVNPKTDLAAFGLMKDIPADYNNIRDLGSRIYTDYLLPFELAAVLLLLGMV
                AAIALVHRKTVNPKRMDPADQVKVRADQGRMRLVKMEAVKPQTESAEESEVSDDLKPK
                EEGKA"
RBS             complement(3874. .3879)
                /gene="nuoJ"
misc_feature    complement(4034. .4525)
                /gene="nuoJ"
                /note="Pfam match to entry PF00499 oxidored_q3,
                NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
                score 77.90, E-value 2e-19"
gene            complement(4570. .5277)
                /gene="NMA0007"
CDS             complement(4570. .5277)
                /gene="NMA0007"
                /note="NMA0007, unknown, len: 235 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0007"
                /protein_id="CAB83327.1"
                /db_xref="GI:7378785"
                /db_xref="SPTREMBL:Q9JX89"

  Query Match              2.8%;  Score 37.6;  DB 1;  Length 340806;
  Best Local Similarity   52.6%;  Pred. No. 2e+02;
  Matches   82;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     870 cgaattggttgacacagacaaccgcgtgatttccggttcggtattgaacggcgcgattac 929
           ||||||||||  |     | | |||    |  ||      | | | |||| |||||| |
Db  230044 CGAATTGGTTTCCGTCCTCGAACGCCCTGTAACCAAAGAAGAAATCAACGCCGCGATGAA 229985

Qy     930 acaaggcgcgcacgattatttgggacgctaccacaatcagatttccgttatcgaagaagg 989
           |   | |||   ||| | ||  ||   | || |  |||| |||      |  ||| |  |
Db  229984 AGCCGCCGCCAGCGAATCTTACGGCTACAACGAAGATCAAATTGTTTCTTCCGACGTTGT 229925

Qy     990 ccgcagcaaagagctgttcggctgggttgcgccgca 1025
           | ||| | || |    |    ||  | |||| | ||
Db  229924 CGGCATCGAATACGGTTCGCTCTTCGATGCGACCCA 229889


RESULT  40
AF324244
LOCUS       AF324244               4499 bp    DNA     linear   PLN 24-JUL-2001
DEFINITION  Phaseolus vulgaris translation initiation factor 2 (IF2) gene,
```

```
          complete cds; nuclear gene for chloroplast product.
ACCESSION   AF324244
VERSION     AF324244.1  GI:12958749
KEYWORDS    .
SOURCE      Phaseolus vulgaris.
  ORGANISM  Phaseolus vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1  (bases 1 to 4499)
  AUTHORS   Campos,F., Garcia-Gomez,B.I., Solorzano,R.M., Salazar,E.,
            Estevez,J., Leon,P., Alvarez-Buylla,E.R. and Covarrubias,A.A.
  TITLE     A cDNA for Nuclear-encoded Chloroplast Translational Initiation
            Factor 2 from a Higher Plant Is Able to Complement an infB
            Escherichia coli Null Mutant
  JOURNAL   J. Biol. Chem. 276 (30), 28388-28394 (2001)
   PUBMED   11356831
REFERENCE   2  (bases 1 to 4499)
  AUTHORS   Campos,F., Garcia-Gomez,B., Solorzano,R., Salazar,E., Estevez,J.,
            Leon,P. and Covarrubias,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2000) Biologia Molecular de Plantas, Instituto de
            Biotecnologia, Universidad Nacional Autonoma de Mexico, Apartado
            Postal 510-3, Cuernavaca, Morelos 62250, Mexico
FEATURES             Location/Qualifiers
     source          1. .4499
                     /organism="Phaseolus vulgaris"
                     /db_xref="taxon:3885"
     mRNA            join(<189. .2890,3117. .3309,4020. .>4163)
                     /gene="IF2"
                     /product="translation initiation factor 2"
     gene            <189. .>4163
                     /gene="IF2"
     CDS             join(189. .2890,3117. .3309,4020. .4163)
                     /gene="IF2"
                     /codon_start=1
                     /product="translation initiation factor 2"
                     /protein_id="AAK09431.1"
                     /db_xref="GI:12958750"
                     /translation="MLILVGSKQGTMSSLASPVSLGSLMGVSSSGRSHSGVRRVSFSR
                     GNCKGRKRWHCLSLSVCRYSVTTTDFIADQGNSVSLDSNSNSSSSSKSGGDDGTGFVL
                     KPPPKPVLKAPDNRMTHLGPSRTTGDVEERNKVIESLGEVLEKAEKLGSSKVNGDKNN
                     GSVNKPVRNNANASPRTERPVNSAASLKSKTLKSVWRKGDSVASVQKVVKEVPKPSYN
                     KNEEEKSQTRGGEKVVSQTRAPQPPSKPQPLKPQQPSKPQPALLSKPSIAPPPVKKPV
                     VLRDKGAAETSVKSKEKKSPILIDKFASKKPVVDPLIAQAVLAPPKPGKAPSPGKFKD
                     DFRKKGALAGGGRRRRILDDEDVIQDASELNVSIPGAATARKGRKWSKASRKAARLQA
                     ARDAAPVKVEILEVGDSGMLVEELAYCLATSEGEILGYLYSKGIKPDGVQTIDKDMVK
                     MICKEYDVEVIDADPVKVEGLVKKREILDEDDLDKLKDRPPVITIMGHVDHGKTTLLD
                     YIRKSKVAASEAGGITQGIGAYKVQVPFDGKTLPCVFLDTPGHEAFGAMRARGASVTD
                     IAVIVVAADDGIRSQTNEAIAHAKAAGVPIVIAINKIDKDGANPERVMQELSSIGLMP
                     EDWGGNTPMVPISALKGKNVDDLLETVMLVAELQELKANPDRSAKGTVIEAGLDKSKG
                     PLATFIVQNGSLRRGDIVVCWRSFWKGRALFDDGGKRVDEATPSIPVQVIGLNNVPIA
                     GDVFEVVESLDAARERAETRAESLRNERISAKAGDGKITLSSLASAVSSGKLSGLDLH
                     QLNIILKVDLQGSIEAVRKALQVLPQENVTLKFLLEATGDVNTSDVDLAVASKAIIMG
                     FNAXTPGSVKSYADNKAVEIRLYRVIYELIDDVRKAMEGLLEPVEEQLTIGSAVVRAV
                     FSSGSGRVAGCMVTEGKVLKDCGIRVKRKGKIVHVGIIDSLRRVKEIVKEVNAGLECG
                     LGLEDFDDWEEGDIIEPSTQLRRRGPLKRPQHQWQLLWRE"
BASE COUNT     1273 a    759 c   1135 g   1329 t      3 others
ORIGIN


  Query Match             2.8%;  Score 37.4;  DB 8;  Length 4499;
  Best Local Similarity   48.4%;  Pred. No. 1.8e+02;
  Matches  104;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     134 tcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagacaaaaagaatccgg 193
           |||   ||||   ||  | | | ||||  ||||| |  ||||| | |||||  |  ||
Db    1756 TCACACAAGGAATTGGGGCCTATAAAGTGCAAGTACCTTTTGATGGCAAAACGTTGCCCT 1815

Qy     194 gcgtggtgtttactgcgccggcttcaggcaaaatcgccgcgattcaccgtggcgaaaagc 253
           | ||  |  ||    | || |     |    | | |  |||||      | | | |
Db    1816 GTGTTTTCCTTGACACTCCTGGACATGAGGCATTTGGAGCGATGAGGGCTCGTGGAGCAA 1875

Qy     254 gcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatcgagtttgaacgct 313
           | |||    |    |  || ||||  || |  | | | || | |||    | | ||
Db    1876 GTGTAACAGACATTGCTGTTATTGTTGTGGCTGCCGATGATGGAATTCGTTCTCAAACAA 1935

Qy     314 acgcacctgaagcgctggcaaacttaagcggcgaa 348
           | | | ||   || |  || ||    |   || | |
Db    1936 ATGAAGCTATTGCTCATGCCAAAGCAGCGGGAGTA 1970


RESULT  41
STMSVIDEOX/c
LOCUS       STMSVIDEOX              7365 bp    DNA     linear   BCT 16-MAY-1996
DEFINITION  Streptomyces violaceoruber Tu22 dTDP-glucose dehydrtatase (graE)
            gene, complete cds, dTDP-glucose synthetase (graD) gene, complete
            cds, ORF4 (Streptomyces coelicolor mmr homolog), complete cds.
ACCESSION   L37334
VERSION     L37334.1  GI:763509
KEYWORDS    dTDP-D-glucose 4,6-dehydratase; dTDP-glucose synthase; transport
            protein.
SOURCE      Streptomyces violaceoruber (clone pOJ446-22-24, strain Tu22) DNA.
  ORGANISM  Streptomyces violaceoruber
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 7365)
  AUTHORS   Bechthold,A., Sohng,J.K., Smith,T.M., Chu,X. and Floss,H.G.
  TITLE     Identification of Streptomyces violaceoruber Tu22 genes involved in
            the biosynthesis of granaticin
  JOURNAL   Mol. Gen. Genet. 248 (5), 610-620 (1995)
  MEDLINE   96027933
FEATURES             Location/Qualifiers
     source          1. .7365
                     /organism="Streptomyces violaceoruber"
                     /strain="Tu22"
                     /db_xref="taxon:1935"
                     /clone="pOJ446-22-24"
     CDS             complement(151. .816)
                     /note="ORF1; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA99938.1"
                     /db_xref="GI:763510"
                     /translation="MLCDSGIGLSPFVILHLHLHHGTEFTMPALRSRTRTAAIAVPLA
                     LAGLMATTSGAQAAPVKTYFDCRVGTVTGLAYSTYVDGVAPAHVRPGATFDVTLDPAP
                     ITPNPAYNKEVKNVALQFKLPAGAKVLSYEVADRLAAGTPELQIDGTKAVLRAAGPYV
                     ANAPFDLPKITIKLKAPTTRGTLETPFGGTSHDDPGFTWTFISPNWTARPSSPAGPTS
                     RSR"
     gene            complement(889. .1854)
                     /gene="graE"
     CDS             complement(889. .1854)
                     /gene="graE"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="dTDP-glucose dehydratase"
                     /protein_id="AAA99939.1"
                     /db_xref="GI:763511"
                     /translation="MRLLVTGAAGFIGSHYVREILAGSYPESDDVHVTVVDRLTYAGR
                     RDNLPEHHERLDFVHGDICDRDLLDRVLPGHDAVVHFAAESHVDRSLTGPGEFVRTNV
                     MGTQQLLDAALHAGVDRVLHVSTDEVYGSLDSGTWTEDSPLLPNSPYAASKASTTWSA
                     APTTVRHGLDVRITRCSNNYGPRQHPEKLIPNFVTRLLTGRQVPLYGDGRNVREWLHV
                     DDHCRALQLVLTKGRAGEIYNIGGGSGMSNREMTARLLDLLGADWDMVRHVEDRLGHD
                     FRYAIDDSKIREELGYAPRWSIESGLGAVVDWYRDHPDFWRAPAS"
     gene            complement(1851. .2909)
                     /gene="graD"
     CDS             complement(1851. .2909)
                     /gene="graD"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-glucose synthetase"
                     /protein_id="AAA99940.1"
                     /db_xref="GI:763512"
```

```
                     /translation="MKALICPEDGDRLRPFTYSMPKQLVPVANKPILVHCLENVRAIG
                     VEEVAVVVGDRADEVRAVVGDGSAFGLDVTYLQQEAPLGLAHCVSIAEEFLGDEDFVM
                     YLGDNILAEGIAESARAFRDERSAARLLLTKVADPRAYGVAETDATGRVHALVEKPER
                     PRSDLAVIGVYFFTAAVHDAVRAIEPSARGELEITDAIQYLVERGDRVVADEYTGYWK
                     DTGSPDDLLDCNRVLLGRLRPGVHGEVDAASTVEGTVVVEAGRSWSDSRLVGPLVVGA
                     GSVVRGSELGPYTALGRDCVLEDAGIRDSIVLDGVSIQGVRGLSGSLIGRSAAVRTGE
                     AAGRRLIIGDHTQAEVAA"
     RBS             complement(1862. .1866)
                     /gene="graD"
                     /note="putative"
     RBS             complement(2922. .2925)
     CDS             complement(3279. .4952)
                     /note="ORF4; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA99941.1"
                     /db_xref="GI:763513"
                     /translation="MTDTANRRSATPDSGADAVDKRRIRLIMTGLLLLGLFMAALDQT
                     IISAALRTIADDLNGLSEQAWANTSYMITSVIMTALYGKLSDIYGRRPVYCTAVGVFV
                     LGSVLCGLAQSMTTLAVFRGQGIGAGGLMSLAFAILTDLVPLAERSRYQAWFGAVFGV
                     SAVVGPVAGGFFAGLDSFLGASGWRWAFFINVPIGVAGALIIAALVRVQPGRAQHKFD
                     IAGLLALIACLLPLLFAVEQGPAWGWTDPGTLVLFGLGAVGLVLFSLAQKRAADAALL
                     PTPMFRNPLFSVYNGVNVLVGAAVFGALSVLPLYLQMVKGLSPTQAGLMMLPQTLGIV
                     VAGRIARPYVTRTGRYKAVLLSGLVLMVVATFWFGVLTADTSLWQTGVAAGVMGFGIG
                     LCWQVMLIAIQTGVAPQYMGAGMGSFTFFRQIGGTGRHRVFLSMFFGAVGEKVAAAYR
                     GAASDPAYTAAVNDPAVTGQAANKVLLGPGRRVGLDNSSFLDHADPRLAKPFLEGLAE
                     AMQTVFVVSGVMLLIALVLAAVPKEKPPRAGRRFAGRRKETGKQPAAKK"
     RBS             5413. .5416
     CDS             5423. .6370
                     /note="ORF5; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA99942.1"
                     /db_xref="GI:763514"
                     /translation="MPLVAATGLPGAPITSRTMFDFMQRDRHGNPLEIPKDPHERNLF
                     NGRGMARLALGSMEGLVPLVERWQPDVLVAGALSYAAPLVAHRFGLPWVRHALNMGEP
                     SIIDLSAAAELAPELEEMGLSAIPEPDMYVEICPPGARRPDAGPAQFMRYVPFNTQRA
                     LEPWMYAKGDRPRVLVSAGSRVTADYEADALSALVEKVAGLDVELLIAAPQEIADALG
                     DLPDNVRAGCVPLDVVLRTCDLLVHRAGGNTMLHAIVCGVPQLVIPAMPKQVGMSARL
                     AEYGAAIMLTAGQDDSPENVPRPAGSCWRTRRTRPGPTS"
BASE COUNT     1098 a   2580 c   2668 g   1019 t
ORIGIN


  Query Match             2.8%;  Score 37.4;  DB 1;  Length 7365;
  Best Local Similarity   55.9%;  Pred. No. 1.9e+02;
  Matches   71;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy   1090 ttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtacttacgagcgc 1149
           |||| | | |||||| ||| || || ||| || || | |||||  |       ||| |
Db   2393 TTCACCGCCGCCGTCCACGACGCCGTCCGGGCGATCGAGCCGAGCGCCCGCGGCGAACTG 2334

Qy   1150 gtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtcggcgatacc 1209
           | |||  ||   |  |||| |  |  | ||       ||||     ||||||| |||
Db   2333 GAGATCACCGACGCGATCCAGTACCTCGTGGAGCGCGGCGACCGGGTCGTCGCCGACGAG 2274

Qy   1210 gacagcg 1216
            ||| ||
Db   2273 TACACCG 2267


RESULT  42
SVI011500
LOCUS       SVI011500              39250 bp    DNA     linear   BCT 24-MAR-1999
DEFINITION  Streptomyces violaceoruber Tu22 granaticin biosynthetic gene
            cluster.
ACCESSION   AJ011500
VERSION     AJ011500.1  GI:4218538
KEYWORDS    acyl carrier protein; aromatase;
            CDP-4-keto-6-deoxyglucose-3-dehydratase(E1); cyclase;
            cyclase-dehydratase; disulphide bond-forming protein;
            dTDP-1-glucose synthase; dTDP-glucose-4,6-dehydratase;
            dTDP4-keto-6-deoxyhexose reductase; FMN-dependent monooxygenase;
            FMN:NADH oxidoreductase; glycosyl transferase; gra-orf 1-37 genes;
            hydroxylase; pathway-specific transcriptional activator; polyketide
            ketoreductase; polyketide synthase chain length factor; polyketide
            synthase keto-acyl synthase; serine-threonine protein kinase;
            transcriptional activator; transmembrane protein; tRNA-Ala gene;
            tRNA-Alanine; two-component response regulator; two-component
            sensory kinase.
SOURCE      Streptomyces violaceoruber.
  ORGANISM  Streptomyces violaceoruber
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 39250)
  AUTHORS   Sherman,D.H., Malpartida,F., Bibb,M.J., Kieser,H.M., Bibb,M.J. and
            Hopwood,D.A.
  TITLE     Structure and deduced function of the granaticin-producing
            polyketide synthase gene cluster of Streptomyces violaceoruber Tu22
  JOURNAL   EMBO J. 8 (9), 2717-2725 (1989)
  MEDLINE   90060034
REFERENCE   2  (bases 1 to 39250)
  AUTHORS   Bechthold,A., Sohng,J.K., Smith,T.M., Chu,X. and Floss,H.G.
  TITLE     Identification of Streptomyces violaceoruber Tu22 genes involved in
            the biosynthesis of granaticin
  JOURNAL   Mol. Gen. Genet. 248 (5), 610-620 (1995)
  MEDLINE   96027933
REFERENCE   3  (bases 1 to 39250)
  AUTHORS   Ichinose,K., Bedford,D.J., Tornus,D., Bechthold,A., Bibb,M.J.,
            Revill,W.P., Floss,H.G. and Hopwood,D.A.
  TITLE     The granaticin biosynthetic gene cluster of Streptomyces
            violaceoruber Tu22: sequence analysis and expression in a
            heterologous host
  JOURNAL   Chem. Biol. 5 (11), 647-659 (1998)
  MEDLINE   99051446
REFERENCE   4  (bases 1 to 39250)
  AUTHORS   Hopwood,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) Hopwood D.A., Genetics, John Innes Centre,
            Norwich Research Park Colney Norwich, Norfolk NR4 7UH, U.K
FEATURES             Location/Qualifiers
     source          1. .39250
                     /organism="Streptomyces violaceoruber"
                     /strain="Tu22"
                     /db_xref="taxon:1935"
     gene            complement(1099. .2417)
                     /gene="gra-orf7"
     CDS             complement(1099. .2400)
                     /gene="gra-orf7"
                     /codon_start=1
                     /transl_table=11
                     /product="putative serine-threonine protein kinase"
                     /protein_id="CAA09628.1"
                     /db_xref="GI:4218539"
                     /translation="MGEVFAGRYELIDPIGRGGVGAVWRAWDHRRRRYVAAKVLQQSD
                     AHTLLRFVREQALRIDHPHVLAPASWAADDDKVLFTMDLVAGGSLAHVIGDYGPLPPR
                     FVCTLLDQLLSGLAAVHAEGVVHRDIKPANILLDVTGTGRPHLRLSDFGISMRKGEPR
                     LTETNYVVGTPGYFAPEQMMGAEPDFPADLFAVGLVALYLLQGRKPDSRALIEYFLAH
                     GTPGAPQGIPEPLWQVLAGLLQPDPQARFKTATGVRKALTAAVEMLPEAGPGEEPVEV
                     FDQIGPCPPASARTARAPSPAAPAPVRVQRHPEEPQQPQQPLPTLSETGSFHLPPPPR
                     QQPAPPPQPVAAPAPHPAAVTPLAAHTDPALAYVPTAAVPYEQASTRSYTAHAPRIPA
                     QASPAAQSPAQRPGPPPKVAIPVLFVALICFAVGIWALTQT"
     RBS             complement(2414. .2417)
                     /gene="gra-orf7"
     gene            complement(2474. .3031)
                     /gene="gra-orf8"
     CDS             complement(2474. .3022)
                     /gene="gra-orf8"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA09629.1"
                     /db_xref="GI:4218540"
                     /translation="MDAAQQEATARARELQRSWYGEPLGALFRRLIDDLGLNQARLAA
                     VLGLSAPMLSQLMSGQRAKIGNPAVVQRVQALQELASQVADGSVSAAEATDRMDEIKK
                     SQGGSVLTGTSQTTNSSGAPTVRRVVREIQSLLRSVAAAGDIIDAADSLAPAHPELAE
```

```
                FLRVYGAGRTGDAVAHYESHQN"
RBS             complement(3028. .3031)
                /gene="gra-orf8"
tRNA            3152. .3226
                /gene="tRNA-ala"
                /product="tRNA-Ala"
gene            3152. .3226
                /gene="tRNA-ala"
gene            3232. .3267
                /gene="tRNA ala"
terminator      3232. .3267
                /gene="tRNA ala"
                /note="putative terminator (stem 16 bp, loop 4 bp)"
protein_bind    complement(3290. .3346)
                /gene="gra-orf8"
                /note="5 CGA repeats"
                /bound_moiety="SARP"
gene            complement(3290. .3346)
                /gene="gra-orf8"
RBS             3454. .3457
                /gene="gra-orf9"
gene            3454. .4313
                /gene="gra-orf9"
CDS             3462. .4313
                /gene="gra-orf9"
                /codon_start=1
                /transl_table=11
                /product="pathway-specific transcriptional activator"
                /protein_id="CAA09630.1"
                /db_xref="GI:4218541"
                /translation="MDVGILGPLAVRLNGQSIVPSAGKPRQLLALLAIRAERVVTVPT
                LMEEIWGDAIPKSAATTLQTYILHLRRKVTAALRAGQPGAAADRAGTGAGAGDAKSVL
                STCFGGYRLALRPEDSDLREFQSLAARGSVALEAGDARAASAMLERALGLWRGTALVD
                VPVGRVLGVEVLGMEEQRTRVLEQRIEADLLLGRHASLLAELRMLVARHPMNENLAAQ
                LMTALYRSGGVWRALEVFQELRRTLIEELGVEPSHRLQLLHRAVLSGELELDPRHAGL
                ELMRTGG"
terminator      4381. .4413
                /gene="gra-orf9/gra-orf10"
                /note="putative bidirectional terminator (stem 12 bp, loop
                9 bp)"
gene            4381. .4413
                /gene="gra-orf9/gra-orf10"
gene            complement(join(4418. .5038,5051. .5140))
                /gene="gra-orf10"
CDS             complement(4418. .5026)
                /gene="gra-orf10"
                /codon_start=1
                /transl_table=11
                /product="putative two-component response regulator"
                /protein_id="CAA09631.1"
                /db_xref="GI:4218542"
                /translation="MIRILLAEDMEMVRGALVALLGLEDDLEVVAQVERGDKILPEAL
                KVRPDVAIIDIGLPGIDGIKAAALLAEELPSCRTLILTSLSRPGVLRRALDAKVAGFL
                PKDASPQDLAEAVRRIAAGHRAIDPQLALAAWDSAESPLTEREMEVLRLTAEGDEAPE
                IAARLHLSTGTVRNYLTSVVTKLNARNRVDAVRIAHESGWLV"
RBS             complement(5036. .5038)
                /gene="gra-orf10"
protein_bind    complement(5051. .5140)
                /gene="gra-orf10"
                /note="4 CC/AA repeats"
                /bound_moiety="SARP"
gene            complement(join(5125. .6345,6567. .6602))
                /gene="gra-orf11"
CDS             complement(5125. .6345)
                /gene="gra-orf11"
                /codon_start=1
                /transl_table=11
                /product="putative two-component sensory kinase"
                /protein_id="CAA09632.1"
                /db_xref="GI:4218543"
                /translation="MNSREQQFDELDSSASYLSHEDSDLPAPRLARTVLLAAMAGYSL
                VTVVNILGAGLSPAETCVGFTALLVVFGLQVAHSRPQARQSPYWRRVTTLTLQAVFCF
                VPLALFHMAWGAMGGFLAGSVLLLLPGRAAWPLYGAIGASLFAYSAAAGHTLLDSVYT
                LEATLVTGLVVFGLSSLTGLVTRLHAHRADLARLAVAGERLRFSRDLHDLLGYNLSAI
                TLKTELVRRLVPDMPERARDETDSILALSRQSLADVRKVATGYRDMSLREEAAATRSM
                LEAAGIEVEAAIDIGPLPSQIDTVLATVLREGVTNALRHSRVHHCTIRCWEKDGVVTL
                SLGNDGVDPEDGGYEGVDRDARRADDPGGTGLGNLTYRLTRVGGSLTTELSDNGWFRM
                TARAPVHPHVCTTR"
protein_bind    complement(6567. .6602)
                /gene="gra-orf11"
                /note="2 CC/GA repeats"
                /bound_moiety="SARP"
gene            complement(6688. .8981)
                /gene="gra-orf12"
terminator      complement(6688. .6711)
                /gene="gra-orf12"
                /note="putative terminator (stem 8 bp, loop 8 bp)"
CDS             complement(6716. .8971)
                /gene="gra-orf12"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA09633.1"
                /db_xref="GI:4218544"
                /translation="MTGEAVYLVSGRGVRAVRGAPAEALAPLLDGTRTPDEVLREAAR
                ARDARRPLPRGVRRALAELWSAGLLRAHPDPAATPGTGAGPPAAHDRSAEAFWDLAGT
                TPPARTAVRLLVLDGPVSEEEARAACAASGLDVGRPGAQASDAEVTLVLCTDYLDPRL
                ADVDALQRAAGRPWILARPGGPEAWTGPVLRPDDGPCWHCLAERLRLQRGSEAPLHRA
                LGCTSPVARPDASLAASRAAGLHTAVLETAKWLAGVPGPTRDAVHTVDALTLDSARHP
                VARRPQCPSCGDAGLVRARVRAPFVPAPRPKAADGGNGHRALSPREVLERYGHLVGPV
                TGVVREVRRDPRSPAFTHSFLSGHNLAMRAPTLAGHSAGLRALSGGKGLTSEEARVSA
                LCEAVERYSGSRHGDEPVVRDSFRAFGPAAVHPDTVRLYDPRQLRDRDAWNASGSPFQ
                YVGRPFDEDAPVDWTPVWSLTGGVQRLLPTSLLYFDADPAAPLRADSNGNAAGSSPED
                AFVQGFLELVERDAVALWWYNRTRAPACDLDAFAASGDAWIEPLRAGYRSLGREIWAL
                DLTSDFGIPVVAALSRRTDGPPEDVLFGFGAHLDPRIAVRRALTEMGQLLPAVVDARP
                DGSGYTVDDPHALSWWRTATVAGQPHLLPDPQRAPSRPGSWTYTARADLSEDVAALAA
                LADERGLEVLVLDQTRPDIGLPVVKVVVPGMRQFWPRFAPGRLFDVPVALGRLAEPTP
                YERLNPLPLFV"
RBS             complement(8978. .8981)
                /gene="gra-orf12"
gene            complement(9059. .9245)
                /gene="gra-orf13"
CDS             complement(9059. .9232)
                /gene="gra-orf13"
                /function="unknown"
                /codon_start=1
                /transl_table=11
```

```
  Query Match              2.8%;  Score 37.4;  DB 1;  Length 39250;
  Best Local Similarity   55.9%;  Pred. No. 2e+02;
  Matches   71;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy   1090 ttcaacacagccgtcaacggcggcgaccgcgccatggtgccgattggtacttacgagcgc 1149
          |||| | | |||||| ||| || || ||| || || | |||||  |      ||| |
Db  13431 TTCACCGCCCGCCGTCCACGACGCCGTCCGGGCGATCGAGCCGAGCGCCCGCGGCGAACTG 13490

Qy   1150 gtgatgcccttggatatcctgcccaccctgcttttgcgcgatttaatcgtcggcgatacc 1209
          | |||  ||   |  |||| |  |  | ||       ||||     |||||| |||
Db  13491 GAGATCACCGACGCGATCCAGTACCTCGTGGAGCGCGGCGACCGGGTCGTCGCCGACGAG 13550

Qy   1210 gacagcg 1216
            ||| ||
Db  13551 TACACCG 13557


RESULT  43
AF429315/c
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN


  Query Match             2.8%;  Score 37.4;  DB 9;  Length 125020;
  Best Local Similarity   9.3%;  Pred. No. 2.2e+02;
  Matches   53;  Conservative  274;  Mismatches  243;  Indels   1;  Gaps    1;

Qy     306 tgaacgctacgcacctgaagcgctggcaaacttaagcggcgaagaagtgcgccgcaacct 365
           |: :| | :  |:  :|:: |: |   :   : ::  :::  : : : :  :::| :
Db   51173 TSCMCCCRRGCCMKKKGMRWCSTTYTGMSGGKSRWWTSSMACMYWYKGSRRKTKSMATSG 51114

Qy     366 gatccaatccggtttgtggactgcgctgcgcacccgtccgttcagcaaaattcctgccgt 425
            : : |: : ||  : :: :::|::  : :    |: :::  : |: :  ::|::   |:
Db   51113 CMRMGAMRSKGGMRKYWSCRYKGMRWGRWSGYKGCSGYMSMGMYGSGRTSKYCSY-TGGY 51055

Qy     426 cgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggctgccga 485
           |:  :    :: :|  :|| ::::  : ::::  |    ::|    ::|  :  || :
Db   51054 CSKCSAKMCKSKSTSKSCCTKSKYSCSMGRYSSYCTCWGGRMCKGMKSSCCASRSGCKSY 50995

Qy     486 ccctacggtcattatcaaagaagccgccgaggatttcaaacgcggcctgttggtattgag 545
            :   :  :  : |: :: :::::: : ::      : |   :|:: ::: : : :::
Db   50994 KMAGSWRCRGRRGMSCWYWSSMKSYSMASYKSSARGRGMKCYKTSCMKYYYSMHBKSSHD 50935

Qy     546 ccgtttgaccgaacgcaaaatccatgtttgtaaggcagctggcgcagacgtgccgtctga 605
           ::|::  : :     |:: : :|: :: :  ::  :::: :::  :  :::: :: :
Db   50934 BSGYKSYRKSYBYWTCMMKDSMCWSVDSHMMRDMMYDKMRMKVSSGTRGYHBHSAWMCKS 50875

Qy     606 aaatgctgccaacatcgaaacacatgaattcggcggcccgcatcctgccggtttgagtgg 665
           : :::  |:: :::  : : :   : : :  ::::::::|: ::|:::::   :::::|
Db   50874 MSRDSDSGYSGMYHMKSYWKSSASKHMYYSGMKSSVYBMDCMSMYTSBSKSVRMBRRWSG 50815

Qy     666 cacgcacattcatttcatcgagccggtcggcgcgaataaaaccgtgtggaccatcaatta 725
           : ::::    :: :::: : :::: |  ::| :::   ::::: |: ::: :  |:   :
Db   50814 SYBRYRKCAVMRRYBHRSSTRSVMYGGKRSCWVRDSVSRRRSVHTSMSSWRMCRCRGARM 50755

Qy     726 tcaagatgtaattaccattggccgtttgtttgcaacaggccgtctgaacaccgagcgcgt 785
              :   : | :|:: :  : ::    ::| : |:: :| |   :: :| ::|  ::
Db   50754 AKSCAGKMSCVMTGMCVHSRCGSMKKCWSCDDCCHRABDAHGVCRCBWYKBCADBCYKRD 50695

Qy     786 gattgccctaggtggttctcaagtcaacaaaccgcgcctcttgcgtaccgttttgggtgc 845
           ::   : : :   ::::::: ::  :::::||::: | ::  :: :: ::: ::  :
Db   50694 SVSCWSAHVWYWVBBHBSKSSRBGWRHMMHACHKSMCKKSVVKVWHVKMKKCYWCCRSCW 50635

Qy     846 gaaagtatcgcaaattactgcgggcgaattg 876
           :  :| :  : :::::  : ||: :: :::
Db   50634 BCSWGVHRGRDRVVKKRTARGGGRARMYKYR 50604


RESULT  44
AC104429
LOCUS       AC104429              154330 bp    DNA     linear   HTG 11-DEC-2001
DEFINITION  Oryza sativa chromosome 3 clone OSJNBb0006P09, *** SEQUENCING IN
            PROGRESS ***, 8 ordered pieces.
ACCESSION   AC104429
VERSION     AC104429.1  GI:17488564
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 154330)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
            Saski,C., Henry,D., Oates,R. and Simmons,J.
  TITLE     Rice Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154330)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
            Saski,C., Henry,D., Oates,R. and Simmons,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2001) Clemson University Genomics Institute,
            Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1      1981: contig of 1981 bp in length
            *                     gap of unknown length
            *     1982     32672: contig of 30691 bp in length
            *                     gap of unknown length
            *    32673     87321: contig of 54649 bp in length
            *                     gap of unknown length
            *    87322     94650: contig of 7329 bp in length
            *                     gap of unknown length
            *    94651     96734: contig of 2084 bp in length
            *                     gap of unknown length
            *    96735     97792: contig of 1058 bp in length
            *                     gap of unknown length
            *    97793    139678: contig of 41886 bp in length
            *                     gap of unknown length
            *   139679    154330: contig of 14652 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154330
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="3"
                     /clone="OSJNBb0006P09"
BASE COUNT    43409 a  33401 c  33796 g  43714 t     10 others
```

```
ORIGIN


  Query Match             2.8%;  Score 37.4;  DB 2;  Length 154330;
  Best Local Similarity   51.5%;  Pred. No. 2.2e+02;
  Matches   86;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     955 cgctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgg 1014
           | | ||  ||| |  ||   | | ||  | ||   |  || ||  |||  |  |  |
Db  135662 CTCGACAGCAAGCCCATCCTCTTCATGAACGAGTTCAACACCATCGAGGAGCCCAACGAC 135721

Qy    1015 gttgcgccgcagccggacaaatactccatcacgcgtacaaccctcggccatttcctgaaa 1074
           |  ||||||| |||| |||  |||     || ||  |      || || | |||| |
Db  135722 GCGGCGCCGCTGCCGACCAAGTACGTGGCCAAGCTCAAGCAGATCCGCGAGTTCCCGGGC 135781

Qy    1075 aacaaactcttcaagttcaacacagccgtcaacggcggcgaccgcgc 1121
           |||     | |||||| |  ||  | | || |  ||   |  | |||
Db  135782 AACGCCGACCTCAAGTACGGCATCGGCCTCGAGAGCCACTTCGCCGC 135828


RESULT  45
AY033407
LOCUS       AY033407               3867 bp    DNA     linear   BCT 07-MAY-2001
DEFINITION  Myxococcus xanthus USC7-1p and USC7-2p genes, complete cds.
ACCESSION   AY033407
VERSION     AY033407.1  GI:13959815
KEYWORDS    .
SOURCE      Myxococcus xanthus.
  ORGANISM  Myxococcus xanthus
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
REFERENCE   1  (bases 1 to 3867)
  AUTHORS   Poza,M., Sieiro,C. and Villa,T.G.
  TITLE     Clone USC7 from Myxococcus xanthus CECT 422 strain
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3867)
  AUTHORS   Poza,M., Sieiro,C. and Villa,T.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Microbiology, University of Santiago de
            Compostela (Faculty of Pharmacy), Campus Sur, Santiago de
            Compostela, A Coruna 15782, Spain
FEATURES             Location/Qualifiers
     source          1. .3867
                     /organism="Myxococcus xanthus"
                     /strain="CECT 422"
                     /db_xref="taxon:34"
                     /clone="USC7"
     CDS             5. .511
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="USC7-1p"
                     /protein_id="AAK49012.1"
                     /db_xref="GI:13959816"
                     /translation="MLELMYAAGLRVSELVLLPATAVNLRQGVLRVTGKGSKERLVPL
                     GEESQHWLERYLQQSRPQLVGKGRVHALADGQTPLFIEPTLHALTRQAFWHLVKRHAQ
                     VAGIDPARISPHGLRHSFATHLLNRGADLRALQMLLGHSSLSTTQIYTLVAREHLQKL
                     HARHHPRG"
     CDS             1451. .2545
                     /note="similar to chitinases"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="USC7-2p"
                     /protein_id="AAK49013.1"
                     /db_xref="GI:13959817"
                     /translation="MWQPSMARLYRFEAGDACRIPGFFSLLQAITDRLGQRALRIVTA
                     ELPDERAAQALGLHVTHLAGDVGRRVEAPQRAAVVDPGGTGGVVGIARPRQVLQHLPV
                     LDAGFIRARSGAAGGLVQAVVGIGHAGPATPVIADAQVELGGGHATADEGLDGIVAVV
                     QLLVAGGRIIRRIEQRRLVIGLRIPRAVEVVGHHVDVVQVRVVRGRIGDADLVDAGTH
                     RHRQQQARTHRIELAAELGHQRGEVGVVFRLAAFDAAGHRVFPVDVDAVEYAQRGTGA
                     TGTVGHRQVALDVGVDARGDEGLAVFRARRAGEAARPGPATQRDQHLHARVLGLELLE
                     LVPVAAQRLVPGIGAAADTLGGTEGLGVVG"

BASE COUNT      524 a   1210 c   1265 g    643 t    225 others
ORIGIN


  Query Match             2.8%;  Score 37.2;  DB 1;  Length 3867;
  Best Local Similarity   17.2%;  Pred. No. 2e+02;
  Matches   45;  Conservative  102;  Mismatches  115;  Indels    0;  Gaps    0;

Qy     319 cctgaagcgctggcaaacttaagcggcgaagaagtgcgccgcaacctgatccaatccggt 378
           |: ::: ::: :::  :|: : |:: ::| ::   :  : :::  :   ::   :|::
Db     895 CMGKMRCMRYSSKMCKRCWCMCGSRYYRASSRCCCRGMYWRMWGTSAASGYYKKCYCRRC 954

Qy     379 ttgtggactgcgctgcgcacccgtccgttcagcaaaattcctgccgtcgatgccgagccg 438
           ::: |: : |  :::| : || : ::   : :::    ::: ::|::    |:|| |:::
Db     955 YWSSGKCSSGRCYYKCTSSCCRRSSYCMMMSRYRSCSMKYSCRYCSYGTYCGYCGCGYYS 1014

Qy     439 ttcgccatcttcgtcaatgcgatggacaccaatccgctggctgccgaccctacggtcatt 498
             :|:|: :: |  :|      |:  :|:| :  :  ::: :|:| :| :  :  :  :|
Db    1015 MMSGSCRASKMCMSSAYSWGCGTSWYSASCGRCRYCRWSKGKGSCYRCKSSCSCWWKSWT 1074

Qy     499 atcaaagaagccgccgaggatttcaaacgcggcctgttggtattgagccgtttgaccgaa 558
             ::::::  :|:::|::     | :  :::|| :    |::  : ::|    :  :
Db    1075 CRSMRRSWYAYCSMYGWRCKMMMCCRYWRSSGCAKCRCCGWMSRKYRMCYRGMRKTMCCG 1134

Qy     559 cgcaaaatccatgtttgtaagg 580
           ||||    ||  |   | ||||
Db    1135 CGCATGGGCCTGGGCAGCAAGG 1156


Search completed: June 30, 2002, 06:19:59
Job time: 47017 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 06:19:59 ; Search time 27321.5 Seconds
(without alignments)
1029.420 Million cell updates/sec

Title: US-09-303-518D-127
Perfect score: 1344
Sequence: 1 atgattaaaatcaaaaaagg.....ccnttgagaaggaaggctga 1344

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	
	1	1327	98.7	1344	6	A96094	A96094 Sequence 12
c	2	1327	98.7	311321	1	NMA3Z2491	AL162754 Neisseria
	3	1229.4	91.5	1344	6	A96092	A96092 Sequence 12
c	4	1229.4	91.5	12273	1	AE002412	AE002412 Neisseria
c	5	1229.4	91.5	349980	6	AX044029	AX044029 Sequence
	6	1165.4	86.7	1344	6	A96098	A96098 Sequence 13
	7	562	41.8	1558	1	APU24492	U24492 Actinobacil
	8	554.2	41.2	10563	1	AE006171	AE006171 Pasteurel
	9	530	39.4	13299	1	U32702	U32702 Haemophilus
	10	506.2	37.7	6048	1	AF165980	AF165980 Vibrio ha
	11	503	37.4	5916	1	AF117331	AF117331 Vibrio ch
c	12	503	37.4	13382	1	AE004300	AE004300 Vibrio ch
	13	477.4	35.5	5078	1	VANQRBOL	Z37111 V.alginolyt
	14	475.8	35.4	6908	1	AB008030	AB008030 Vibrio al
c	15	473	35.2	16903	1	AE004724	AE004724 Pseudomon
	16	431.6	32.1	474	6	A96090	A96090 Sequence 12
c	17	333.6	24.8	96109	6	AX067460	AX067460 Sequence
	18	122.2	9.1	1356	1	AF144076	AF144076 Porphyrom
c	19	101.4	7.5	4926	1	HIU20229	U20229 Haemophilus
c	20	45.6	3.4	125020	9	AF429315	AF429315 Homo sapi
c	21	44.6	3.3	36200	1	BACGNTZA	D78193 Bacillus su
c	22	44.6	3.3	215534	1	BSUB0021	Z99124 Bacillus su
c	23	41.2	3.1	1641	1	AB015974	AB015974 Pseudomon
	24	41	3.1	628	8	AF316898	AF316898 Haberlea
c	25	40.6	3.0	11375	1	AE002001	AE002001 Deinococc
c	26	39.2	2.9	340806	1	NMA1Z2491	AL162752 Neisseria
	27	38.2	2.8	2557	3	AY071018	AY071018 Drosophil
	28	38.2	2.8	65140	6	AX211705	AX211705 Sequence
	29	38.2	2.8	89791	3	AC005268	AC005268 Drosophil
c	30	38.2	2.8	101085	2	AC018341	AC018341 Drosophil
	31	38.2	2.8	123580	1	AF263912	AF263912 Streptomy
	32	38.2	2.8	125401	6	AX211739	AX211739 Sequence
	33	38.2	2.8	164035	3	AC007888	AC007888 Drosophil
c	34	38.2	2.8	305020	3	AE003452	AE003452 Drosophil
	35	38	2.8	69283	2	AC110159	AC110159 Mus muscu
	36	38	2.8	125020	9	AF429315	AF429315 Homo sapi
	37	37.2	2.8	9743	14	AF435865	AF435865 Rubella v
	38	37.2	2.8	9743	14	AF435866	AF435866 Rubella v
	39	37.2	2.8	9762	6	AX009468	AX009468 Sequence
	40	37.2	2.8	9762	14	AF188704	AF188704 Rubella v
	41	37.2	2.8	9762	14	RUBNOPR	L78917 Rubella vir
c	42	37.2	2.8	112721	8	AC016780	AC016780 Genomic S
c	43	37.2	2.8	142737	8	AC027658	AC027658 Oryza sat
	44	37.2	2.8	329861	1	NMA5Z2491	AL162756 Neisseria
c	45	37	2.8	763	6	AX019436	AX019436 Sequence
c	46	37	2.8	1598	6	AX019435	AX019435 Sequence
c	47	37	2.8	2042	1	AF031148	AF031148 Methyloco
	48	37	2.8	9807	1	AE009707	AE009707 Brucella
	49	36.8	2.7	5931	1	BCU97042	U97042 Burkholderi
c	50	36.8	2.7	10695	1	AE007265	AE007265 Sinorhizo
	51	36.6	2.7	9762	14	AB047329	AB047329 Rubella v
	52	36.6	2.7	9762	14	AB047330	AB047330 Rubella v
c	53	36.6	2.7	10029	1	AE008364	AE008364 Agrobacte
	54	36.6	2.7	10051	1	AE009249	AE009249 Agrobacte
	55	36.6	2.7	10486	1	AE001656	AE001656 Chlamydia
c	56	36.6	2.7	15680	1	AE002164	AE002164 Chlamydop
	57	36.6	2.7	42348	1	SC27G11	AL391763 Streptomy
c	58	36.6	2.7	65979	8	AF254799	AF254799 Hordeum v
c	59	36.6	2.7	189050	1	AL646066	AL646066 Ralstonia
	60	36.6	2.7	300550	1	AP002547	AP002547 Chlamydop
	61	36.4	2.7	1325	6	AX168124	AX168124 Sequence
	62	36.4	2.7	38634	1	SC26G5	AL359152 Streptomy
c	63	36.4	2.7	188050	1	AL646072	AL646072 Ralstonia
	64	36.2	2.7	4081	14	AF274291	AF274291 Culex nig
	65	36.2	2.7	4623	1	AF419157	AF419157 Aeromonas
c	66	36.2	2.7	11525	1	AE005110	AE005110 Halobacte
	67	36.2	2.7	108252	14	AF403738	AF403738 Culex nig
c	68	36	2.7	762	6	AX019439	AX019439 Sequence
	69	36	2.7	3230	9	AK023114	AK023114 Homo sapi
c	70	36	2.7	10944	1	AE002563	AE002563 Neisseria
c	71	36	2.7	11162	1	AE005655	AE005655 Escherich
c	72	36	2.7	32039	1	SC7A1	AL034447 Streptomy
c	73	36	2.7	172325	6	AX044035	AX044035 Sequence

c	74	36	2.7	243184	1	AP002568	AP002568 Escherich
	75	35.8	2.7	601	1	AF170251	AF170251 Streptomy
c	76	35.8	2.7	14029	1	AE007175	AE007175 Mycobacte
c	77	35.8	2.7	121125	1	MTV025	AL022121 Mycobacte
	78	35.8	2.7	154330	2	AC104429	AC104429 Oryza sat
	79	35.8	2.7	195767	1	NMA7Z2491	AL162758 Neisseria
	80	35.6	2.6	960	1	PSU83882	U83882 Pseudomonas
	81	35.6	2.6	1462	1	PSECATEX	L10655 Sphingomona
	82	35.6	2.6	3867	1	AY033407	AY033407 Myxococcu
	83	35.6	2.6	4004	1	SSPZ84817	Z84817 Sphingomona
	84	35.6	2.6	6600	14	RVM33NP	X72393 Rubella vir
	85	35.6	2.6	9755	14	RUBCG	M15240 Rubella vir
	86	35.6	2.6	9759	6	I63470	I63470 Sequence 1
	87	35.6	2.6	27000	1	SCE33	AL355774 Streptomy
	88	35.4	2.6	690	3	AF069057	AF069057 Giardia i
c	89	35.4	2.6	11274	1	AE004603	AE004603 Pseudomon
	90	35.2	2.6	1918	1	AB008137	AB008137 Enterobac
	91	35.2	2.6	1918	1	AB008142	AB008142 Enterobac
c	92	35.2	2.6	3949	10	MMSEMF	X97817 M.musculus
	93	35	2.6	296	6	AX260544	AX260544 Sequence
	94	35	2.6	372	1	LDEXCA	Z25772 L.discophor
	95	35	2.6	1935	6	E15828	E15828 Nitrosomona
	96	35	2.6	2033	1	AF010470	AF010470 Haloferax
	97	35	2.6	3379	10	AB015422	AB015422 Mus muscu
	98	35	2.6	4776	1	AB018706	AB018706 Nitrosomo
c	99	34.8	2.6	34214	14	ADRGENOME	L19443 Human adeno
	100	34.8	2.6	43961	1	SCM1	AL133422 Streptomy

ALIGNMENTS

RESULT 1
A96094
LOCUS A96094 1344 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 127 from Patent WO9924578.
ACCESSION A96094
VERSION A96094.1 GI:6779939
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1344)
AUTHORS Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
TITLE Neisserial antigens
JOURNAL Patent: WO 9924578-A 127 20-MAY-1999;
PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES Location/Qualifiers
source 1. .1344
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 323 a 347 c 368 g 289 t 17 others
ORIGIN

Query Match 98.7%; Score 1327; DB 6; Length 1344;
Best Local Similarity 100.0%; Pred. No. 9.3e-313;
Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
Db       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 60

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
Db      61 TATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
Db     121 CCCTNGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
Db     181 AAAAAGNATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCAT 240

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
Db     241 CGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy     301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
Db     301 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGAANTNNGNNGC 360

Qy     361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
Db     361 AATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCT 420

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
Db     421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGCG 480

Qy     481 gcagaccctgtggttgtgatcaaagaagccgngcgangatttcagacgangtntgctggta 540
Db     481 GCAGACCCTGTGGTTGTGATCAAAGAAGCCGNGCGANGATTTCAGACGANGTNTGCTGGTA 540

Qy     541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
Db     541 TTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 600

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
Db     601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTG 660

Qy     661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
Db     661 AGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATC 720

Qy     721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
Db     721 AATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 780

Qy     781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
Db     781 CGCGTGATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTG 840

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
Db     841 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
Db     901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 960

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db     961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy    1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
Db    1021 CCGCAGCCGGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAA 1080

Qy    1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
Db    1081 CTCTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACT 1140

Qy    1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
Db    1141 TACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 1200

Qy    1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
Db    1201 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
```

```
Db   1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCCCGCTGTTGCGTAAGGTGCTG 1320

Qy   1321 gaaaccnttgagaaggaaggctga 1344
          |||||||||||||||||||||||
Db   1321 GAAACCNTTGAGAAGGAAGGCTGA 1344


RESULT   2
NMA3Z2491/c
LOCUS       NMA3Z2491              311321 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 3/7.
ACCESSION   AL162754 AL157959
VERSION     AL162754.2  GI:7379424
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 311321)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 311321)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .311321
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     RBS             115. .120
     gene            129. .587
                     /gene="NMA0698"
     CDS             129. .587
                     /gene="NMA0698"
                     /note="NMA0698, possible periplasmic protein, len: 152 aa;
                     unknown, contains a probable N-terminal signal sequence
                     and lies within a region of unusually low GC content"
                     /codon_start=1
                     /transl_table=11
                     /product="putative periplasmic protein".
                     /protein_id="CAB83985.1"
                     /db_xref="GI:7379425"
                     /db_xref="SPTREMBL:Q9JQQ2"
                     /translation="MLTKLKILLFLFLFVFVLAINLLFFFFSSDIESFGNYQFEYVYD
                     KGWPANYILVMKDGNEGNFDKIISGLVLEYYKEDDNIYFSYIDGQGFASDSCYYKPEI
                     LYGKIILNKNHIININSMEKNNFLSEDKIMKGTRNWLADPKNKCNIQTLD"
     gene            742. .1077
                     /gene="NMA0699"
     CDS             742. .1077
                     /gene="NMA0699"
                     /note="NMA0699, pseudogene, probable ABC transporter
                     protein, len: 336 bp; similar to C-termini of many ABC
                     transporters e.g. TR:Q46973 (EMBL:U47048), mtfB,
                     Escherichia coli microcin transport protein (707 aa),
                     fasta scores; E(): 2.6e-19, 53.2% identity in 111 aa
                     overlap. Contains PS00211 ABC transporters family
                     signature. NMA0686 may be the remainder of this
                     pseudogene"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="putative ABC transporter protein (pseudogene)"
     misc_feature    868. .912
                     /gene="NMA0699"
                     /note="PS00211 ABC transporters family signature"
     gene            complement(1101. .2327)
                     /gene="NMA0700"
     CDS             complement(1101. .2327)
                     /gene="NMA0700"
                     /note="NMA0700, possible ribonuclease BN, len: 408 aa;
                     similar to SW:RBN_ECOLI (EMBL:L19201), rbn, Escherichia
                     coli ribonuclease BN (EC 3.1.-.-), fasta scores; E():
                     2.3e-28, 34.3% identity in 268 aa overlap. Longer than rbn
                     at the C-terminus. Also similar to TR:O85449
                     (EMBL:AF067083) Vitreoscilla sp. hypothetical protein (376
                     aa), fasta scores; E(): 0, 45.0% identity in 333 aa
                     overlap. Contains a region similar to NMA0677, fasta
                     scores; E(): 4.3e-08, 64.4% identity in 45 aa overlap.
                     Contains hydrophobic, possible membrane-spanning regions,
                     as does rbn"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribonuclease BN"
                     /protein_id="CAB83987.1"
                     /db_xref="GI:7379426"
                     /db_xref="SPTREMBL:Q9JVT9"
                     /translation="MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLL
                     ALVPVLTVMVAVASIFPVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTA
                     IGSVMLVVTSLMLIRTIDNTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFM
                     VGSVQDAALASGAPQWSGALRTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATA
                     FCLETARSLFTWYMGNFDGYRSIYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQ
                     GEAFRRGFDSRGRFDDVLKILLLLDAAQKEGKALPVQEFRRHINMGYDELGELLEKLA
                     RHGYIYSGRQGWVLKTGADSIELNELFKLFVYRPLPVERDHVNQAVDAVMMPCLQTLN
                     MTLAEFDAQAKKQQQS"
     misc_feature    1376. .1385
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(2417. .2426)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2444. .2453
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            2486. .2743
                     /gene="NMA0701"
     CDS             2486. .2743
                     /gene="NMA0701"
                     /note="NMA0701, possible pseudogene, len: 258 bp; shows
                     weak similarity to part of SW:WRBA_ECOLI (EMBL:M99166),
                     wrbA, Escherichia coli Trp repressor binding protein (197
                     aa), fasta scores; E(): 0.064, 29.3% identity in 82 aa
                     overlap. Also similar to part of TR:O85450 (EMBL:AF067083)
                     Vitreoscilla sp. Trp repressor binding protein (fragment)
                     (124 aa), fasta scores; E(): 6.6e-14, 56.6% identity in 83
                     aa overlap"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="hypothetical protein NMA0701 (pseudogene)"
     misc_feature    complement(2755. .2764)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2778. .2787
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             2861. .2864
     gene            2871. .3530
                     /gene="NMA0702"
     CDS             2871. .3530
```

```
                /gene="NMA0702"
                /note="NMA0702, len: 219 aa; similar to TR:Q52925
                (EMBL:Z50189), exsB, Rhizobium meliloti putative regulator
                of succinoglycan biosynthesis (not a transcriptional
                regulator) (234 aa), fasta scores; E(): 3.1e-15, 32.7%
                identity in 205 aa overlap. Also similar to many
                hbacterial ypothetical proteins e.g. SW:YBAX_HAEIN
                (EMBL:U32798), HI1191, Haemophilus influenzae hypothetical
                protein (196 aa), fasta scores; E(): 0, 79.2% identity in
                173 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0702"
                /protein_id="CAB83989.1"
                /db_xref="GI:7379427"
                /db_xref="SPTREMBL:Q9JVT8"
                /translation="MSNQKALVIFSGGQDSTTCLIQAIQTYGRENVQAITFQYGQRHA
                VELERARWIAQDLGVKQTVLDLSLMRQITHNALMDDTAAIETAENGVPNTFVDGRNAL
                FLLYAAIYAKGQGIRHIIAGVCETDFSGYPDCRDVFVKSMNVTLNLAMDYDFQIHTPL
                MYLTKAQTWALADEMGALDYIREQTHTCYNGIVGGCRECPSCILRERGLAEYLESKKA
                V"
     misc_feature    3522. .3531
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
     RBS             3547. .3550
     gene            3558. .4076
                /gene="NMA0703"
     CDS             3558. .4076
                /gene="NMA0703"
                /note="NMA0703, len: 172 aa; unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein NMA0703"
                /protein_id="CAB83990.1"
                /db_xref="GI:7379428"
                /db_xref="SPTREMBL:Q9JVT7"
                /translation="MPKLHMFYLGGNAGRSNIEVHDIQFAVCDDYREAVPALKAAWFG
                DADKIHIDGWQVVEWADGYDIAVSETPKTKMPSENAPRLYFANVGGYRAGQLAEAHAF
                GLFAAATPAEAKQKALQTLLTDSYVQQHKDNLKDVDNLLALDRIGNFHIRLTPNPHGK
                PAEIGFQGYLPI"
     misc_feature    3782. .3791
                /gene="NMA0703"
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
     gene            4083. .4505
                /gene="NMA0704"
     misc_feature    4083. .4502
                /gene="NMA0704"
                /note="Pfam match to entry PF01242 PTPS, 6-pyruvoyl
                tetrahydropterin synthase, score 10.80, E-value 1.7e-06"
     CDS             4083. .4505
                /gene="NMA0704"
                /note="NMA0704, len: 140 aa; similar to many hypothetical
                proteins e.g. SW:YB90_HAEIN (EMBL:U32798), HI1190,
                Haemophilus influenzae hypothetical protein (141 aa),
                fasta scores; E(): 5.1e-32, 58.0% identity in 138 aa
                overlap. Shows very weak similarity to eukaryotic
                6-pyruvoyl-tetrahydropterin synthases e.g. SW:PTPS_RAT
                (EMBL:M77850), pts, Rattus norvegicus 6-pyruvoyl
                tetrahydrobiopterin synthase precursor (EC 4.6.1.10) (144
                aa), fasta scores; E(): 0.0062, 27.3% identity in 154 aa
                overlap. Contains Pfam match to entry PF01242 PTPS,
                6-pyruvoyl tetrahydropterin synthase"

  Query Match             98.7%;  Score 1327;  DB 1;  Length 311321;
  Best Local Similarity   98.7%;  Pred. No. 1e-312;
  Matches 1327;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55476 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 55417

Qy     61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55416 TATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 55357

Qy    121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55356 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 55297

Qy    181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
          |||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  55296 AAAAAGAATCCGGGCGTGGTGTTTACCGCGCCGGTTTCAGGCAAAATCGCCGCCATCCAT 55237

Qy    241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55236 CGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 55177

Qy    301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
          |||||||||||||||||||||||||||||||||||||||||||||||| ||| |  |  ||
Db  55176 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 55117

Qy    361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  55116 AATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTACCCGTCCGTTCAGCAAAATCCCT 55057

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  55056 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCG 54997

Qy    481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
          |||||||||||||||||||||||||||||||| ||| |||||||||||| || ||||||||
Db  54996 GCAGACCCTGTGGTTGTGATCAAAGAAGCCGCCGAGGATTTCAGACGAGGTTTGCTGGTA 54937

Qy    541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54936 TTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 54877

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54876 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTG 54817

Qy    661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54816 AGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATC 54757

Qy    721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54756 AATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 54697

Qy    781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54696 CGCGTGATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTG 54637

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54636 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 54577

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54576 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 54517

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54516 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 54457

Qy   1021 ccgcagccggacaaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54456 CCGCAGCCGGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAA 54397

Qy   1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54396 CTCTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACT 54337

Qy   1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  54336 TACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 54277

Qy   1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54276 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 54217

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  54216 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGTAAGGTGCTG 54157

Qy   1321 gaaaccnttgagaaggaaggctga 1344
          |||||| ||||||||||||||||||
Db  54156 GAAACCATTGAGAAGGAAGGCTGA 54133


RESULT   3
A96092
LOCUS       A96092                  1344 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 125 from Patent WO9924578.
ACCESSION   A96092
VERSION     A96092.1  GI:6779938
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1344)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 125 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1344
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      329 a    366 c    365 g    284 t
ORIGIN


  Query Match             91.5%;  Score 1229.4;  DB 6;  Length 1344;
  Best Local Similarity   94.2%;  Pred. No. 5.8e-289;
  Matches 1266;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | ||
Db      1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 60

Qy     61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
          || ||||| || | ||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy    121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy    181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
          |||||| |||||||||||||||||| ||||| | |||||||||||||||||| || ||
Db    181 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 240

Qy    241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
          || |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy    301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
          ||||| |||||||||| || |||||| |||||||||||||||||| ||| |  |  ||
Db    301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 360

Qy    361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
          || |||||||||||||||||||||||||||||||| | |||||||||||||||||| |||
Db    361 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 420

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
```

```
Db    421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 480

Qy    481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
          || |||||  |||  | |||||||||||||| ||| ||||||| |||  |  || |||||
Db    481 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 540

Qy    541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
          |||||||||||||||| || |||||||||| |||||||||||||||||||||||||||||
Db    541 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 600

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 660

Qy    661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
          |||||||||||||||||||||| |||||||||| || || |||||||| |||||||||
Db    661 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 720

Qy    721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
          |||||||||||||||||| |||| || ||||||||||||||||||||||||||||||||||
Db    721 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 780

Qy    781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
          |||||||||||  | |||||||||||||||||||||| ||||||||||||||||||||
Db    781 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    841 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 900

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 960

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy   1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1021 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 1080

Qy   1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
          |||||||||||||  |||||||||||||| ||||||||||||||||||||||||||||||
Db   1081 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 1140

Qy   1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
          |||||||||| |||||  | || |||||||| |||||||||||||||||||||||||||||
Db   1141 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 1200

Qy   1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
          |||||||||||||||||||||||||||||||||||| |||||||||||||||| || ||||||
Db   1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 1320

Qy   1321 gaaaccnttgagaaggaaggctga 1344
          |||||| ||||||||||||||||||
Db   1321 GAAACCATTGAGAAGGAAGGCTGA 1344


RESULT   4
AE002412/c
LOCUS       AE002412                12273 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 54 of 206 of
            the complete genome.
ACCESSION   AE002412 AE002098
VERSION     AE002412.1  GI:7225783
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
```

```
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 12273)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 12273)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12273
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            97. .435
                     /gene="NMB0557"
     CDS             97. .435
                     /gene="NMB0557"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40985.1"
                     /db_xref="GI:7225784"
                     /translation="MSDESPIIFTDSCCAKVADLIAEENNPDLKLRVFVNGGGCSGFQ
                     YGFTFDEIKNDDDFEIEKNGLVFLVDPMSYQYLVGAEIDYTESLQGSQFVIRNPNAET
                     TCGCGSSFSV"
     gene            595. .1059
                     /gene="NMB0558"
     CDS             595. .1059
                     /gene="NMB0558"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40986.1"
                     /db_xref="GI:7225785"
                     /translation="MIEMSKDYRNDLYDVYVSYPPQVDRGLIRECLKENLGEEKAEGL
                     IESLDSKPQVLVEEKCTWAKREELHDYFSYLGLDIITRRYMELETVVPPEEGEGEGEG
                     ADGEMPEYLELHGGREDDISAPSQPEPPSRNIKLLVFGLLIAFLGYLLGKIF"
     gene            1094. .2605
                     /gene="NMB0559"
     CDS             1094. .2605
                     /gene="NMB0559"
                     /note="similar to GB:M87049 SP:P27853 PID:148233 GB:U00096
                     PID:2367309 percent identity: 74.73; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ubiquinone biosynthesis protein AarF"
                     /protein_id="AAF40987.1"
                     /db_xref="GI:7225786"
                     /translation="MKWLKRLTVIVGTFYRYRLAGLCASLMGSGWICALLKMMPQSSK
                     LKNEPPAVRLRLALESLGPIFIKFGQVLSTRPDLIPHDYAVELAKLQDKVPPFDARLS
                     REQIEKSLGQSIEKLYAEFETEPIASASIAQVHKARLHSGEQVAVKVLRPNLLPVIEQ
                     DLSLMRFGAGWVERLFADGKRLKPREVVAEFDKYLHDELDLMREAANASQLGRNFQNS
                     DMLIVPKVFYDYCTSDVLTIEWMDGTPVSDIAKLKADGIDLHKLADYGVEIFFTQVFR
                     DGFFHADMHPGNILVAADNRYIALDFGIVGTLTDYDKRYLAINFLAFFNRDYRRVATA
                     HIESGWVPADTRAEELEAAVRAVCEPVFNKPISQISFGLVLMRLFEVSRRFNVEIQPQ
                     LVLLQKTLLNIEGLGRQLDPDLDLWKTAKPFLVKWMNGQVGPKALWRNLKNEAPDWAQ
                     IIPSLPRKISALIDENRQQEMRDAYIHLVKVQQRQSLWLAVIAVVLLLILLLK"
     gene            complement(2962. .3780)
                     /gene="NMB0560"
     CDS             complement(2962. .3780)
                     /gene="NMB0560"
                     /note="similar to GB:L42023 SP:P43886 PID:1005429
                     PID:1220683 PID:1204857 percent identity: 72.97;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="serine acetyltransferase"
                     /protein_id="AAF40988.1"
                     /db_xref="GI:7225787"
                     /translation="MKKDHLNTTGFDLWHTIREETAAAVSAEPMLASFLHQTVLRHES
                     LGSVLAYHLSSKLGSPIMDVRALFEIYQQALGSDTQIGKCVEADLKAIYERDPACDEY
                     SLPLLYFKGFHAIQAHRINHRLYLDGRKTLAYFLQNRMSEVFGVDIHPAARFGYGLML
                     DHATGFVAGETAVLGNNISILHGVTLGGSGKEGGDRHPKIGDGVMIGANASILGNIRI
                     GSNAKIGAGSVVVSDVPPSITVVGVPAKPVARSLKTPSADMDQNIQFTEIDFMI"
     gene            3964. .4542
                     /gene="NMB0561"
     CDS             3964. .4542
                     /gene="NMB0561"
                     /note="similar to GP:3777486 percent identity: 61.03;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="grpE protein"
                     /protein_id="AAF40989.1"
                     /db_xref="GI:7225788"
                     /translation="MSEQTQQQNSEEAVENVEAVETVETVGNADGVQEQAAAEPAYED
                     LQARIAELEAQLKDEQLRALANEQNLRRRHQQEIADTHKFAGQKFAVEMLPVKDYLEM
                     ALLDQSGNFDALKMGVQMTLNELQKAFDATQIKEINPKAGDKLDPNIHQAMQAVASEQ
                     EPNTVVGVMKKGYTLSDRVLRPAMVTVAQKEA"
     gene            complement(4819. .5034)
                     /gene="NMB0562"
     CDS             complement(4819. .5034)
                     /gene="NMB0562"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40990.1"
                     /db_xref="GI:7225789"
                     /translation="MKTLLLTFGIFLTVIIGMAVGYIFSKRSIKGSCGGITALGMKKM
                     CDCDTPCDTLQKKLDEEKQAGGIRVDR"
     gene            complement(5060. .6115)
                     /gene="NMB0563"
     CDS             complement(5060. .6115)
                     /gene="NMB0563"
                     /note="similar to GB:L42023 SP:P44550 PID:1003244
                     PID:1222087 PID:1204428 percent identity: 74.70;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="thiamine biosynthesis lipoprotein ApbE"
                     /protein_id="AAF40991.1"
                     /db_xref="GI:7225790"
                     /translation="MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTT
                     YTVKYLSNNRDKLPSPAEIQKRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRIS
                     SDFAHVTAEAVRLNRLTHGALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGI
                     DKIILKQGKDYASLSKTHPKAYLDLSSIAKGFGVDKVAGELEKYGIQNYLVEIGGELH
```

```
                GKGKNARGEPWRIGIEQPNIVQGGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHI
                INPNNKRPISHNLASISVVADSAMTADGLSTGLFVLGETEALKLAEREKLAVFLIVRD
                KGGYRTAMSSEFEKLLR"
     gene       complement(6268. .7485)
                /gene="NMB0564"
     CDS        complement(6268. .7485)
                /gene="NMB0564"
                /note="similar to GB:L42023 PID:1003242 PID:1222086
                PID:1204427 PID:1573127 percent identity: 92.82;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="Na(+)-translocating NADH-quinone reductase,
                subunit F"
                /protein_id="AAF40992.1"
                /db_xref="GI:7225791"
                /translation="MEIILGIVMFTVIVLVLALMILFAKSKLVSEGDITIKVNGEKEL
                TMPAGGKLLGALANEGIFIPSACGGGGSCGQCRVVVKSGGGDILPTELSHISKREARE
                GCRLSCQVNVKTDMDIEVPEEVFGVKKWECTVISNDNKATFIKELKLAIPEGEEVPFR
                AGGYIQIEAPPHTVAYKDFDIPKEYHEDWDKYNLWQYVSKVDEPILRAYSMASYPEEK
                GIIMLNVRIATPPPRVPDAPPGQMSSYIWSLKPGDKVTISGPFGEFFAKDTDAEMVFI
                GGGAGMAPMRSHIFDQLKRLNSKRKITFWYGARSKREMFYVEDFDQLAAEFPNFTWHV
                ALSDPLPEDNWDGYTGFIHNVVYENHLKNHEAPEDCEFYMCGPPIMNQSVIKMLKDLG
                VEDENILLDDFGG"
     gene       complement(7499. .8092)
                /gene="NMB0565"
     CDS        complement(7499. .8092)
                /gene="NMB0565"
                /note="similar to PID:1573126 percent identity: 91.92;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="Na(+)-translocating NADH-quinone reductase,
                subunit E"
                /protein_id="AAF40993.1"
                /db_xref="GI:7225792"
                /translation="MEHYLSLFIKSVFIENMALSFFLGMCTFLAVSKKVSTAFGLGVA
                VIFVLGLSVPVNQLVYSLLKDGAIAEGVDLTFLKFITFIGVIAALVQILEMFLDKFVP
                ALYNALGIYLPLITVNCAIFGAVSFMAQREYNFGESVVYGFGAGLGWMLAIVALAGIT
                EKMKYSDAPKGLKGLGITFIAAGLMAMAFMSFSGIQL"
     gene       complement(8096. .8722)
                /gene="NMB0566"
     CDS        complement(8096. .8722)
                /gene="NMB0566"
                /note="similar to GB:L42023 SP:P43958 PID:1003237
                PID:1222083 PID:1204425 percent identity: 89.42;
```

Query Match 91.5%; Score 1229.4; DB 1; Length 12273;
Best Local Similarity 94.2%; Pred. No. 6e-289;
Matches 1266; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   12069 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 12010

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           || ||||| || | ||||||||||||||||||||||||||||||||||||||||||||||
Db   12009 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 11950

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11949 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 11890

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
           |||||| |||||||||||||||||| ||||| | |||||||||||||||||| || ||
Db   11889 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 11830

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
           || |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   11829 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 11770

Qy     301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
           ||||| |||||||||||| || |||||| |||||||||||||||||| ||| |  |  ||

Db   11769 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 11710

Qy     361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
           || |||||||||||||||||||||||||||||||| | ||||||||||||||||| |||
Db   11709 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 11650

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   11649 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 11590

Qy     481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
           || |||||| ||| | |||||||||||||| ||| ||||||| ||| | || |||||
Db   11589 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 11530

Qy     541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
           |||||||||||||||| || |||||||||| ||||||||||||||||||||||||||||
Db   11529 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 11470

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   11469 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 11410

Qy     661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
           |||||||||||||||||||||| |||||||||| || || |||||||| |||||||||
Db   11409 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 11350

Qy     721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
           |||||||||||||||||| |||| || ||||||||||||||||||||||||||||||||
Db   11349 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 11290

Qy     781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
           ||||||||||  | |||||||||||||||||||||| ||||||||||||||||||||||
Db   11289 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 11230

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   11229 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 11170

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11169 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 11110

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11109 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 11050

Qy    1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   11049 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 10990

Qy    1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
           |||||||||||||  |||||||||||||| ||||||||||||||||||||||||||||||
Db   10989 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 10930

Qy    1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
           |||||||||| |||||  | || |||||||| ||||||||||||||||||||||||||||
Db   10929 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 10870

Qy    1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   10869 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 10810

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   10809 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 10750

Qy    1321 gaaaccnttgagaaggaaggctga 1344
           |||||| ||||||||||||||||||
Db   10749 GAAACCATTGAGAAGGAAGGCTGA 10726
```

RESULT 5
AX044029/c

```
LOCUS       AX044029              349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 108 from Patent WO0066791.
ACCESSION   AX044029
VERSION     AX044029.1  GI:11342913
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 108 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.-seq 1: 1 to
                     349980 349980 bases-seq 108: 300001 to 649980 349980
                     bases-seq 109: 600001 to 949980 349980 bases-seq 110:
                     900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
                     349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
                     113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    84410 a  84863 c  94187 g  86520 t
ORIGIN
```

Query Match 91.5%; Score 1229.4; DB 6; Length 349980;
Best Local Similarity 94.2%; Pred. No. 6.3e-289;
Matches 1266; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| | ||
Db 297909 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 297850

Qy     61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
          || ||||| || | ||||||||||||||||||||||||||||||||||||||||||||||
Db 297849 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 297790

Qy    121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297789 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 297730

Qy    181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
          |||||| |||||||||||||||||| ||||| | ||||||||||||||||| || ||
Db 297729 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 297670

Qy    241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
          || |||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db 297669 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 297610

Qy    301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
          ||||| |||||||||| || |||||| |||||||||||||||||| ||| |  |  ||
Db 297609 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 297550

Qy    361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
          || |||||||||||||||||||||||||||||| | ||||||||||||||||| |||
Db 297549 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 297490

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 297489 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 297430

Qy    481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
          || |||||| |||  | |||||||||||||| ||| ||||||| |||  |  || |||||
Db 297429 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 297370

Qy    541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
          |||||||||||||||| || |||||||||| ||||||||||||||||||||||||||||
Db 297369 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 297310

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db 297309 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 297250

Qy    661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
          |||||||||||||||||||||| |||||||||||  || || |||||||| |||||||||
Db 297249 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 297190

Qy    721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
          |||||||||||||||||| |||| || ||||||||||||||||||||||||||||||||||
Db 297189 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 297130

Qy    781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
          |||||||||||  | |||||||||||||||||||||| ||||||||||||||||||||
Db 297129 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 297070

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db 297069 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 297010

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297009 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 296950

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296949 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 296890

Qy   1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 296889 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 296830

Qy   1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
          |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db 296829 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 296770

Qy   1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
          |||||||||||  |||||  | || |||||||| ||||||||||||||||||||||||||||
Db 296769 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 296710

Qy   1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db 296709 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 296650

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
          |||||||||||||||||||||||||||||||||||| |||||||||||||| || ||||||
Db 296649 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 296590

Qy   1321 gaaaccnttgagaaggaaggctga 1344
          |||||| ||||||||||||||||||
Db 296589 GAAACCATTGAGAAGGAAGGCTGA 296566
```

```
RESULT   6
A96098
LOCUS       A96098                 1344 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 131 from Patent WO9924578.
ACCESSION   A96098
VERSION     A96098.1  GI:6779941
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1344)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 131 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1344
```

```
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      334 a    363 c    362 g    285 t
ORIGIN

  Query Match             86.7%;  Score 1165.4;  DB 6;  Length 1344;
  Best Local Similarity   91.2%;  Pred. No. 2.3e-273;
  Matches 1226;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 60

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           |||||||| || | ||||||||||||||||||||||||||||||||| ||| ||||||
Db      61 TATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGC 120

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           |||| ||||||| |||||||||| || ||||||||||||||||||||||||||||||||||
Db     121 CCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
           |||||| |||||||||| || ||||| ||||| | |||||||||||||||||| || ||
Db     181 AAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCAC 240

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
           || |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy     301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
           ||||||||||||||||  || |||||| ||||||| || ||| ||||  || |  |  ||
Db     301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGC 360

Qy     361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
           || ||||| ||||| || || |||||||||| || | ||||||||||||||||||||||
Db     361 AACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCT 420

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 480

Qy     481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
           || |||||| |||  | |||||||||||||| ||| || |||| |||  |  || |||||
Db     481 GCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTA 540

Qy     541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
           ||||||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db     541 TTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
           |||||||||||||||| |||||||||||||||| |||||||||||| ||||| |||
Db     601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTG 660

Qy     661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
           |||||||||||||||||||| ||||| ||||| || || |||||||| |||||||||||
Db     661 AGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATC 720

Qy     721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
           |||||||||| || ||||| ||||||||||||| | |||||||||||||| ||||||
Db     721 AATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAG 780

Qy     781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
           |||||| |||| ||||| ||    |||||||||||||| ||||||||||||||||||||
Db     781 CGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
           |||||||| || || ||| |||| || ||||||||||||||| ||||||||||||||||
Db     841 GGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
           |||||||||||||||||||| |||||| |||||||||||||| ||||||||||||||||||
Db     901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTAC 960

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy    1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
           |||||||||||||||||||||||||||||| || || |||||||||||||| |||||||||
Db    1021 CCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAA 1080

Qy    1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
           |||||||||||||||||||||||||||||| |||||||||||||||| ||||| ||| |||
Db    1081 CTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACT 1140

Qy    1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
           || |||||||||||||| | |||||||||||||| |||||||||||||||||||||||||
Db    1141 TATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200

Qy    1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
           |||||||||||||||||| || ||||||||||||||||||||||||||||||||||||||
Db    1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
           |||||||||||||||||||||||||||||||| |||||||||||||| || ||||||
Db    1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 1320

Qy    1321 gaaaccnttgagaaggaaggctga 1344
           |||||| |||||||||||||||||
Db    1321 GAAACCATTGAGAAGGAAGGCTGA 1344


RESULT   7
APU24492
LOCUS       APU24492                1558 bp    DNA     linear   BCT 23-FEB-1996
DEFINITION  Actinobacillus pleuropneumoniae 48 kDa outer membrane protein
            (aopA) gene, complete cds.
ACCESSION   U24492
VERSION     U24492.1  GI:1185394
KEYWORDS    .
SOURCE      Actinobacillus pleuropneumoniae.
  ORGANISM  Actinobacillus pleuropneumoniae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Actinobacillus.
REFERENCE   1  (bases 1 to 1558)
  AUTHORS   Cruz,W.T., Nedialkov,Y.A., Thacker,B.J. and Mulks,M.H.
  TITLE     Molecular characterization of a common 48-kilodalton outer membrane
            protein of Actinobacillus pleuropneumoniae
  JOURNAL   Infect. Immun. 64 (1), 83-90 (1996)
  MEDLINE   96110919
REFERENCE   2  (bases 1 to 1558)
  AUTHORS   Mulks,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-1995) Martha H. Mulks, Microbiology, Michigan
            State University, 367 Giltner Hall, East Lansing, MI 48824, USA
FEATURES             Location/Qualifiers
     source          1. .1558
                     /organism="Actinobacillus pleuropneumoniae"
                     /strain="ISU158"
                     /serotype="5"
                     /db_xref="taxon:715"
     gene            95. .1444
                     /gene="aopA"
     CDS             95. .1444
                     /gene="aopA"
                     /note="similar to ngrA gene of Vibrio alginolyticus
                     encoding the alpha subunit of a sodium-translocating
                     NADH-uiquinone oxidoreductase, Genbank Accession Number
                     Z37111"
                     /codon_start=1
                     /transl_table=11
                     /product="48 kDa outer membrane protein"
                     /protein_id="AAC43631.1"
                     /db_xref="GI:1185395"
                     /translation="MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMK
```

```
                     VREGDVVKKGQVLFEDKKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITF
                     TRYEAAQLASLSAEQVKQNLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLA
                     ADPEVVLKEYETDFKDGLTVLTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGV
                     HPAGLVGTHIHFVDPVGATKQVWHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNP
                     RLVRTRLGANLSQLTANELNAGENRVISGSVLSGATAAGPVDYLGRYALQVSVLAEGR
                     EKELFGWIMPGSDKFSITRTVLGHFGKKLFNFTTAVHGGERAMVPIGAYERVMPLDII
                     PTLLLRDLAAGDTDSAQNLGCLELDEEDLALCTYVCPGKNNYGPMLRAALEKIEKEG"
     terminator      1462. .1499
BASE COUNT      432 a    299 c    375 g    452 t
ORIGIN


  Query Match           41.8%;  Score 562;  DB 1;  Length 1558;
  Best Local Similarity  64.6%;  Pred. No. 2.7e-126;
  Matches  874;  Conservative    0;  Mismatches  467;  Indels   12;  Gaps    3;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| |||||| |||||  |  |  | || ||||| || | |||||  ||||| ||
Db      95 ATGATTACAATCAAGAAAGGCTTGGATTTACCTATCGCAGGTACACCGGCACAAGTAATC 154

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
             || ||||      | |||  ||||| ||| ||||||||||||||||||  ||||||||
Db     155 CATAACGGCAATACCGTTAATGAAGTTGCGATGCTTGGCGAAGAATATGTGGGTATGCGT 214

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           || |  ||||||||    ||||||||||  || |||||||| || || || ||||||||
Db     215 CCTTCAATGAAAGTTCGTGAAGGCGATGTAGTGAAAAAAGGTCAGGTTCTTTTTGAAGAT 274

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
           |||||| ||||||||||  || ||||| || || |      || |   | |   | ||  |
Db     275 AAAAAGAATCCGGGCGTTGTATTTACTGCTCCTGCAAGCGGTACCGTGGTTACTATTAAC 334

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
           ||||| |||||||| || ||||||||||||||||||    ||||| ||  | ||  ||||
Db     335 CGCGGTGAAAAGCGTGTTCTTCAGTCGGTCGTGATTAAAGTTGAGGGTGATGAGCAAATT 394

Qy     301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
             ||   ||||| |   | |     || ||      ||   | |||  || |
Db     395 ACCTTTACTCGCTATGAAGCTGCGCAATTAGCGTCGCTATCAGCCGAGCAAGTGAAACAA 454

Qy     361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
           ||||| ||| |||| ||||| ||||||||  | ||||  ||||||||||| ||  |.||
Db     455 AATCTTATCGAATCGGGTTTATGGACTGCATTCCGTACTCGTCCGTTCAGTAAGGTTCCC 514

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
           ||  | || ||    ||||   | ||||| |||||||| |||| |||||||||  | |||
Db     515 GCATTGGACGCGATTCCGTCATCAATCTTTGTCAATGCAATGGATACCAATCCGTTAGCG 574

Qy     481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
           || || || |  || ||  | |||||   ||     ||||| | |    || |    ||
Db     575 GCGGATCCGGAAGTGGTTTTAAAAGAGTACGAAACCGATTTTAAAGACGGTTTAACTGTT 634

Qy     541 ttgagccgtttgaccgagcgtaaaa------tccatgtgtgtaaggcagctggcgcagac 594
           || | ||||||     |  || |||      |  |  | ||||| |  || | |  |
Db     635 TTAACCCGTTTATTTAACGGTCAAAAACCGGTTTACTTATGTAAAGACGCGGACAGCAAT 694

Qy     595 gtgccgtctgaaaatgc---tgccaacatcgaaacacatgaattcggcggcccgcatccg 651
             | ||||        ||   ||      |||  ||   |    ||  ||||    |||||
Db     695 ATCCCGTTAAGTCCGGCGATTGAAGGTATCACAATTAAGTCGTTTAGCGGTGTTCATCCT 754

Qy     652 gccggtttgagtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtt 711
           ||||||||    || ||||| || || ||  |||| || || || || | ||||   |||
Db     755 GCCGGTTTAGTCGGTACGCATATCCACTTTGTTGATCCCGTAGGGGCGACCAAACAAGTT 814

Qy     712 tggaccatcaattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctg 771
           |||  | | ||||||||||||||| ||||| |||||     || ||| | ||||||   ||
Db     815 TGGCACTTAAATTATCAAGATGTGATTGCTATCGGTAAATTATTTACTACAGGCGAACTC 874

Qy     772 aacaccgagcgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgt 831
             ||||| ||  | ||| |  | |  ||| | ||||| || || || ||  |  | |||
Db     875 TTTACCGACCGTATTATTTCGCTTGCCGGTCCGCAAGTGAAAAATCCTCGTTTAGTACGT 934

Qy     832 accgttttgggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaac 891
```

```
           ||   | | |||||||||  | || ||| | || ||    || ||   || ||  || |||
Db     935 ACGCGTCTTGGTGCGAATCTTTCCCAATTAACCGCAAATGAGTTAAATGCCGGTGAAAAC 994

Qy     892 cgcgtgatttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttg 951
           || |||||||| |||||||||||||| ||| |||| | |    ||  ||   ||||| ||
Db     995 CGTGTGATTTCAGGTTCGGTATTGAGCGGTGCGACTGCGGCGGGTCCGGTTGATTACTTA 1054

Qy     952 ggacgctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggc 1011
           || || |||       ||  | |||||  | |  ||||| ||    |||||| | |||||
Db    1055 GGCCGTTACGCATTACAAGTATCCGTACTGGCGGAAGGTCGTGAGAAAGAGTTATTCGGT 1114

Qy    1012 tgggttgcgccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctg 1071
           ||| ||  ||||    |||| |||| |||||| || |||||     | |||||  |||
Db    1115 TGGATTATGCCGGGTTCGGATAAATTCTCCATTACCCGTACCGTGTTAGGCCACTTCGGT 1174

Qy    1072 aaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccg 1131
           |||   ||| | || || ||||| || || ||  |||| || || || || ||||||||
Db    1175 AAA---AAATTATTTAACTTCACTACGGCAGTTCACGGCGGTGAACGTGCAATGGTGCCA 1231

Qy    1132 attggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgat 1191
           || ||| | || ||||| |||||||| |||| ||  | |||||  |  |  | || |||
Db    1232 ATCGGTGCGTATGAGCGTGTAATGCCGTTAGATATTATTCCTACGTTATTACTTCGTGAT 1291

Qy    1192 ttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaa 1251
           |||   | ||||||||||||    || |||   || ||||| || ||  |||| ||||||
Db    1292 TTAGCTGCCGGCGATACCGATTCTGCACAAAACTTAGGTTGTTTAGAGCTGGATGAAGAA 1351

Qy    1252 gacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgt 1311
           ||  | || || || |  |  || ||||||| ||| |  | || || ||| |||| |||
Db    1352 GATTTGGCATTATGTACTTATGTTTGCCCGGGTAAAAATAACTACGGTCCGATGTTACGT 1411

Qy    1312 aaggtgctggaaaccnttgagaaggaaggctga 1344
             | | | || |   | ||||||||||| | |
Db    1412 GCTGCGTTAGAGAAGATCGAGAAGGAAGGTTAA 1444


RESULT   8
AE006171
LOCUS       AE006171               10563 bp    DNA     linear   BCT 08-MAR-2001
DEFINITION  Pasteurella multocida PM70 section 138 of 204 of the complete
            genome.
ACCESSION   AE006171 AE004439
VERSION     AE006171.1  GI:12721691
KEYWORDS    .
SOURCE      Pasteurella multocida.
  ORGANISM  Pasteurella multocida
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Pasteurella.
REFERENCE   1  (bases 1 to 10563)
  AUTHORS   May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
            Kapur,V.
  TITLE     Complete genomic sequence of Pasteurella multocida,Pm70
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
  MEDLINE   21145866
REFERENCE   2  (bases 1 to 10563)
  AUTHORS   Zhang,Q. and Kapur,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
            University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
            55108, USA
FEATURES             Location/Qualifiers
     source          1. .10563
                     /organism="Pasteurella multocida"
                     /strain="PM70"
                     /db_xref="taxon:747"
     gene            94. .1605
                     /gene="PM1326"
     CDS             94. .1605
                     /gene="PM1326"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK03410.1"
```

```
                /db_xref="GI:12721692"
                /translation="MWIDVMTQRDCILKLKNIGKSFFDVPVLKNVDLDIYAGEVHCLV
                GENGAGKSTLCKIIAGIYPFEQGEMFYEGKSYFPKTVKEAQSLGIGFIHQELMLVPQL
                SVLENIFLGNEKTVFCGKMDWQTMRHKTQQIISELELDIDPDAKVADLSIAQQQMVEI
                AKAVFSEYKVIIFDEPTSSISRKNTQTLFKIIHQLKEKNVAMIYISHRLDEFQYIADQ
                VTVLRDGIRTGTMAYRDTSPDEIVKLMVGREINFAEYHRHYLPTREVLRVESLTNKHI
                KSSSFSLYEGEILGFAGLVGAGRTELLRAIFGADEASGAIYLNAQRVEIRSPEDAVKQ
                KIGFITEDRKQQGLILGLSIRENITLPILKRFWNGLFLNKKQERLVAEENRQKLRIVS
                HSQDQISQTLSGGNQQKVIIARWLESGVKILFFDEPTRGIDVGAKSEIYDLMRQFTDN
                GGAIVMVSSDLPELITMSDRVIVMRQGGIVREVKDKAEMTEENLMRLMIGVDE"
gene            1598. .2641
                /gene="rbsC_3"
                /note="PM1327"
CDS             1598. .2641
                /gene="rbsC_3"
                /codon_start=1
                /transl_table=11
                /product="RbsC"
                /protein_id="AAK03411.1"
                /db_xref="GI:12721693"
                /translation="MNRYCMTNLRKIFSKLGIGIILLLMIIGMSLSSDVFLSTNNIIN
                ILLQVSIICVISVGMTYVILTGGIDLSVGSIVALSAVCLGVFTHWGMDWLGENPSSFS
                VLMIVISAIIATIFVGILCGYVNGLVIVYGKVTPFITTLGMMGIARGLALTISDGKTI
                YNFPDTLRFLGNGRIALTETFALPVPVIIALLVVLVSFYVLTQTMFGRQIYALGGNRE
                AVRLSGINIEKLEIKAYVINGALAAIGAIILVGRLNAAQPIAGNGYELDAIAATVIGG
                TSLMGGVGSVVSTSIGALIMGVLQNGLTLLNVTSYLQRLIIGLVIILAVFLDQLRRGE
                VSTRRLKRLFFRE"
gene            3187. .4527
                /gene="PM1328"
CDS             3187. .4527
                /gene="PM1328"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03412.1"
                /db_xref="GI:12721694"
                /translation="MITIKKGLNLPISGSPEQVIRDGNAITEVALLGEEYVGMRPSMK
                VREGDVVKKGQVLFEDKKNPGVVFTAPASGTVTAIHRGAKRVLQSVVIKIEGNEQITF
                EKYTTEQLNQLTSEQVRQNLQASGLWTALRTRPFSKVPAVDATPVSIFVNAMDTNPLC
                ADPQVIVQQSAQAFEAGLTVLSRLHEGKVYLCKAANASIPSPSIANLDVKEFAGPHPA
                GLSGTHIHFIDPVSATKFVWYINYQDVIAVGKLFTTGELDVSRVVSLAGPQVKNPRLV
                RTVLGANLSQLTANEVKDGENRVISGSVLSGAKAAGPVDYLGRYALQVSVLEEGREKE
                FLGWIMPGANKYSLSRTVLGHFSKKLFNFTTALNGGERAMVPIGAYERVMPLDIIPTL
                LLRDLAAGDTDSAQALGCLELDEEDLALCTFVCPGKNEYGPLLRQALDKIEKEG"
gene            4530. .5762
                /gene="nqrB"
                /note="PM1329"
CDS             4530. .5762
                /gene="nqrB"
                /codon_start=1
                /transl_table=11
                /product="NqrB"
                /protein_id="AAK03413.1"
                /db_xref="GI:12721695"
                /translation="MGLKHFIEKIEPAFLPGGKYEKWYALFEATATFLYTPGTVTHKA
                SHVRDALDSKRMMVLVWLSLFPAMFYGMYNVGAQAILATDALGTLQQAISGHWQYALS
                HALGADLTLSAGWASKMLLGATYFLPIYLTIFLVGGFWEVLFAMVRKHEINEGFFVTS
                ILLALIVPPTLPLWQAALATTFGVVVAKEIFGGVGKNFMNPALAARAFLFFAYPAQIS
                GDTVWVAADGFSGATALSQWAVGGEAGLKHVVTGQPITWMDAFIGNIPGSIGEVSTLA
                LIIGAAIIVFARIASWRIIAGVMIGMIATSALFNLIGSSTNPLFAMPWYWHFVLGGFA
                LGMFFMATDPVSAAFTNKGKWWYGILIGAMAVLIRVVNPAYPEGMMLAILFANLFAPV
                FDYLVVQANIKRRRARNV"
gene            5755. .6537
                /gene="PM1330"
CDS             5755. .6537
                /gene="PM1330"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03414.1"
                /db_xref="GI:12721696"
                /translation="MFKNKDSVGGTLLVIILLSLACSIIVAGSAVLLKPTQIEQKELD
                KQKNILSVAGLLQPTTKNSEIKTIYANNIEARLVDLNTGDFAPAQPGFDAAKAVKNPA
                ESTALSAEDDVAGIRVRANFAEVYLVKNDAGQVTQVVLPFYGKGLWSMMYGFMSVQPD
                GNTVNGITYYDQGETPGLGGEIENPKWQAQFPGKKLYTADNQVGLYVGKGASANAEHG
                IDAISGSTLTSNGVNNSFKFWLGQKGFGPFLAKLKAGVLNNG"
gene            6530. .7156
                /gene="PM1331"
CDS             6530. .7156
                /gene="PM1331"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03415.1"
                /db_xref="GI:12721697"
                /translation="MADTKKLKGLLLSPVMDNNPIALQILGICSALAVTTKLETAIVM
                TFAVIFVTAFSNLFISLIRNYIPNSVRIIVQMAIIASLVILVDQILRAYAYGLSKQLS
                VFVGLIITNCIVMGRAEAFAMKSQPLESFVDGIGNGLGYGAILVSVAFIRELIGSGKL
                FGMTVFQTIQDGGWYQTNGLFLLAPSAFFIIGFIIWGIRTLKPNQVEK"
gene            7160. .7756
                /gene="nqrE"
                /note="PM1332"
CDS             7160. .7756
                /gene="nqrE"
                /codon_start=1
                /transl_table=11
                /product="NqrE"
                /protein_id="AAK03416.1"
                /db_xref="GI:12721698"
                /translation="MEHYISLFVKSVFIENMALSFFLGMCTFLAVSKKVSTAFGLGIA
                VIVVLGIAVPVNQLVYSFILKDSALVQGIDLSFLNFITFIGVIAALVQILEMVLDKYF
                PALYNALGIFLPLITVNCAIFGGVSFMVQRDYTFVESVVYGIGAGTGWMLAIVALAGI
                TEKMKYADVPAGLRGLGITFITVGLMALGFMSFSGIQL"
gene            7772. .8995
                /gene="nqr6"
                /note="PM1333"
CDS             7772. .8995
                /gene="nqr6"
                /codon_start=1
                /transl_table=11
                /product="Nqr6"
                /protein_id="AAK03417.1"
                /db_xref="GI:12721699"
                /translation="MEITLGIAMFTVIVLALAVIILFAKSKLVNSGDITIEINDDPSK
                AIHLPAGGKLLGALASQGIFVSSACGGGGSCGQCIVKVTEGGGDILPTELSHISKREA
                KEGYRLSCQVNVKNSMKVELPEEVFGVKKWECTVISNDNKATFIKELKLAIPEGEEVP
                FRAGGYIQIEAEPHTVAYKDFDIPEEYHEDWDKYNLWRYVSKVDEHIIRAYSMASYPE
                EKGIIMLNVRIATPPPNNPDAPPGQMSSYIWSLKAGDKVTISGPFGEFFAKETDAEMV
                FVGGGAGMAPMRSHIFDQLKRLKSKRKMSFWYGARSKREMFYVEDFDTLQAENDNFVW
                HVALSDPQPGDNWDGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPIMNASVIKMLKD
                LGVEDENILLDDFGG"
gene            9164. .10204
                /gene="apbE"
                /note="PM1334"
CDS             9164. .10204
                /gene="apbE"
                /codon_start=1
                /transl_table=11
                /product="ApbE"
                /protein_id="AAK03418.1"
                /db_xref="GI:12721700"
                /translation="MKHILNSIIIGFCSLCLLACHKSPEVVTITGKTMGTTYSVKYID
                DGQIKIKPIEMHGQIEEILKEVNNKMSTYLSTSELSQFNQNTQIDTPVEISADLAFVL
                KEAIRLHGVTEGALDVTVGPIVNLWGFGPEKRQPNETLEEQVEKRRAWVGIEKMKLTE
                QGDKFYLAKTVPQLYIDLSSIAKGFGVDKVAAHLATKGIANYLVEIGGEINAKGKNIE
                GKDWQIAIEKPNFDGSRAVSQIIGLKDFSMATSGNYRNYFEENGRRFSHEIDPKTGYP
                IEHRLASITVLADSTMTADGLSTGLYVLGEEKALEVAEKYNLLVYLISKTEKGFEAKM
                SSAFEKLLNEQH"
gene            10214. .10468
                /gene="PM1335"
CDS             10214. .10468
                /gene="PM1335"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03419.1"
```

```
                     /db_xref="GI:12721701"
                     /translation="MQLFLITFGIFLLIIAGMAIGYIIQRKTIAGSCGGISALGLKKV
                     CDCEEPCDNLKAKLEAGDEEAQQEYNAKFAKNEPHFYEVK"
BASE COUNT      3052 a    1783 c    2451 g    3277 t
ORIGIN


  Query Match             41.2%;  Score 554.2;  DB 1;  Length 10563;
  Best Local Similarity   63.3%;  Pred. No. 2.2e-124;
  Matches  851;  Conservative    0;  Mismatches  490;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| || |||||||| | || || || ||  |||| || || ||||| || ||
Db    3187 ATGATTACGATTAAAAAAGGCTTGAATCTTCCTATTTCGGGAAGCCCAGAGCAGGTAATC 3246

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
             ||| ||    |  || || ||||| || |||||||| ||||| |||| || ||||||
Db    3247 CGTGATGGCAATGCTATCACTGAAGTTGCTTTGCTTGGTGAAGAGTATGTGGGAATGCGC 3306

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           || |  ||||| ||    ||||| |||| || ||||||||||||||||| ||||||||
Db    3307 CCTTCAATGAAGGTGCGTGAAGGTGATGTTGTGAAAAAAGGCCAAGTGCTTTTTGAAGAT 3366

Qy .   181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
           |||||| | || || ||||| || || || || |     || |   | | || ||||||
Db    3367 AAAAAGAACCCTGGTGTGGTTTTCACAGCCCCTGCAAGTGGTACTGTAACTGCGATTCAT 3426

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
           || || |  ||||| ||  | || || || || |||    ||||||| || ||  ||||
Db    3427 CGTGGTGCTAAGCGAGTTTTACAATCTGTAGTAATTAAAATTGAAGGTAATGAACAAATT 3486

Qy     301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
             || |||   ||  |  | |||   ||     |  | |     ||  || | |
Db    3487 ACTTTTGAAAAATATACAACAGAACAATTAAACCAACTTACTTCAGAGCAAGTTCGTCAA 3546

Qy     361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
           ||| |       || || || ||||| || || |||| ||||||||||| ||| | |||
Db    3547 AATTTACAAGCTTCTGGATTATGGACCGCACTTCGTACGCGTCCGTTCAGTAAAGTGCCT 3606

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
           ||||||||||||   ||  |  | || ||||| |||||||||| |||||||  |
Db    3607 GCCGTCGATGCCACACCCGTATCGATTTTCGTGAATGCGATGGATACCAATCCATTATGT 3666

Qy     481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
           || || ||   ||  |  |  || || | |   | | |||   |   || ||   |||
Db    3667 GCCGATCCACAAGTGATCGTGCAACAATCTGCGCAAGCTTTTGAAGCGGGTCTGACTGTA 3726

Qy     541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
           || || |||||    ||| | ||| |  |  ||||||| || ||    |  | ||
Db    3727 TTAAGTCGTTTACATGAGGGGAAAGTGTACCTGTGTAAAGCGGCGAATGCGAGCATTCCA 3786

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
             ||    ||||||  | ||     | ||||| |  || || ||||| || || ||
Db    3787 AGCCCAAGCATTGCCAATCTTGACGTGAAAGAATTTGCAGGTCCACATCCAGCTGGGTTA 3846

Qy     661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
           ||||| ||||| ||||| || ||||| || ||  |||||| ||||   ||||||   |||
Db    3847 AGTGGTACGCATATTCACTTTATTGATCCTGTGAGTGCAACCAAATTTGTTTGGTATATC 3906

Qy     721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
           ||||| |||||||| || ||  | ||     || ||| | |||||      |  |
Db    3907 AATTACCAAGATGTGATAGCGGTGGGTAAATTATTTACCACAGGTGAATTAGATGTTTCA 3966

Qy     781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
           |||||  |  | || | |||| | ||||| || || ||| |  |  | ||||||| ||
Db    3967 CGCGTAGTGTCCTTAGCTGGTCCACAAGTGAAAAATCCAAGATTAGTACGTACCGTATTA 4026

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
           ||||| ||  | || ||| | || |||   ||| ||   ||||  || || |||||||||
Db    4027 GGTGCAAACCTTTCCCAATTAACAGCGAATGAAGTGAAAGACGGTGAAAATCGCGTGATT 4086

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
           || ||||| ||||| |  || || |   |   ||| |    |||||||| || || |||

Db    4087 TCAGGTTCAGTATTAAGTGGTGCAAAAGCTGCAGGTCCTGTTGATTATTTAGGTCGTTAC 4146

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
             ||  |||| ||  | |||||||| ||    |||||  |  | || ||| |   |
Db    4147 GCATTGCAAGTTTCAGTACTTGAAGAAGGTCGTGAGAAAGAATTTCTAGGTTGGATCATG 4206

Qy    1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
           |||    || | |||||||| |  | || |||    | || ||||||   |||   |||
Db    4207 CCGGGTGCGAATAAATACTCACTTTCACGCACGGTGTTAGGGCATTTCTCTAAA---AAA 4263

Qy    1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
             | || || || || || ||  | || || || || || || || ||||| || || ||| ||
Db    4264 TTATTTAATTTTACCACCGCACTTAATGGCGGTGAGCGTGCGATGGTACCTATCGGTGCT 4323

Qy    1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
           || ||||| || |||||  | || ||  | || ||  |  |  |||| |||||    |
Db    4324 TATGAGCGTGTGATGCCATTGGATATTATTCCAACATTATTACTGCGTGATTTGGCAGCA 4383

Qy    1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
           ||||||||||    || || || || || || || || || |  | || ||||||||  | ||
Db    4384 GGCGATACCGATTCTGCACAGGCGTTAGGGTGTTTAGAGCTAGATGAAGAAGATTTAGCC 4443

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
             | || | ||| || ||||| || ||| | ||||| || ||||| || ||| | |   |
Db    4444 CTTTGTACCTTTGTTTGCCCTGGTAAAAATGAATATGGTCCGCTCTTACGTCAAGCTTTA 4503

Qy    1321 gaaaccnttgagaaggaaggctga 1344
           || |   | ||||||||||| | |
Db    4504 GATAAGATCGAGAAGGAAGGTTAA 4527


RESULT   9
U32702
LOCUS       U32702                 13299 bp    DNA     linear   BCT 29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 17 of 163 of the complete genome.
ACCESSION   U32702 L42023
VERSION     U32702.1  GI:1573118
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
  ORGANISM  Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 13299)
  AUTHORS   Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
REFERENCE   2  (bases 1 to 13299)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
REFERENCE   3  (bases 1 to 13299)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 13299)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
```

```
  REMARK     The H. influenzae sequence has been updated by R. Fleischmann. New
             database matches have been assigned, product names have been
             improved, and a number of frame shifts have been corrected. We
             gratefully acknowledge the work of Tatusov et. al. We have
             incorporated their annotation into the /notes fields of the
             corresponding H. influenzae genes
REFERENCE    5  (bases 1 to 13299)
  AUTHORS    White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
             Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
             Medical Center Dr, Rockville, MD 20850, USA
  REMARK     The whole genome was shifted by 588 nucleotides for a new start
COMMENT      On Sep 30, 1996 this sequence version replaced gi:1220812.
FEATURES             Location/Qualifiers
     source          1. .13299
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            complement(85. .1455)
                     /gene="HI0161"
     CDS             complement(85. .1455)
                     /gene="HI0161"
                     /note="similar to SP:P06715 GB:M13141 PID:146248
                     PID:466637 GB:U00096 percent identity: 74.17; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glutathione reductase (gor)"
                     /protein_id="AAC21833.1"
                     /db_xref="GI:1573119"
                     /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                     VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                     LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                     DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                     LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                     LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                     KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                     QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                     EFVTMR"
     gene            complement(1671. .2270)
                     /gene="HI0162"
     CDS             complement(1671. .2270)
                     /gene="HI0162"
                     /note="similar to GB:U00096 PID:1773118 PID:1786637
                     percent identity: 25.82; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="lipoprotein, putative"
                     /protein_id="AAC21834.1"
                     /db_xref="GI:1573120"
                     /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                     VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                     QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                     VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
     gene            2364. .2675
                     /gene="HI0163"
     CDS             2364. .2675
                     /gene="HI0163"
                     /note="similar to PID:663269 percent identity: 46.53;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, putative (bolA)"
                     /protein_id="AAC21835.1"
                     /db_xref="GI:1573121"
                     /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                     SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                     R"
     gene            2966. .4309
                     /gene="HI0164"
     CDS             2966. .4309
                     /gene="HI0164"
                     /note="similar to PID:663269 PID:663270 percent identity:
                     64.96; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreducatase subunit A, Na
                     translocating"
                     /protein_id="AAC21836.1"
                     /db_xref="GI:1573122"
                     /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                     VREGDVVKKGQVLFEDKKNPGVIFTAPASGTITAINRGEKRVLQSVVINVEGDEKITF
                     AKYSTEQLNTLSSEQVKQNLIESGLWTALRTRPFSKVPSIESEASSIFVNAMDTNPLA
                     ADPSVVLKEYSQDFTNGLTVLSRLFPSKPLHLCKAGDSNIPTADLENLQIHDFTGVHP
                     AGLVGTHIHFIDPVGIQKTVWHINYQDVIAVGKLFTTGELYSERVISLAGPQVKEPRL
                     VRTTIGANLSQLTQNELSAGKNRVISGSVLCGQIAKDSHDYLGRYALQVSVIAEGNEK
                     EFFGWIMPQANKYSVTRTVLGHFSKKLFNFTTSENGGERAMVPIGSYERVMPLDILPT
                     LLLRDLIVGDTDGAQELGCLELDEEDLALCSFVCPGKYEYGSILRQVLDKIEKEG"
gene            4312. .5547
                     /gene="HI0166"
CDS             4312. .5547
                     /gene="HI0166"
                     /note="similar to PID:663269 PID:663271 percent identity:
                     71.33; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, subunit B
                     (nqrB)"
                     /protein_id="AAC21837.1"
                     /db_xref="GI:1573123"
                     /translation="MGLKNLFEKMEPAFLPGGKYSKLYPIFESIYTLLYTPGTVTHKN
                     THVRDALDSKRMMITVFLALFPAIFYGMYNVGNQAIPALNQLGNLDQLIANDWHYALA
                     SSLGLDLTANATWGSKMALGAIFFLPIYLVVFTVCTIWELLFSVVRGHEVNEGMFVST
                     ILFALIVPPTLPLWQAALGITFGIIVAKEIFGGVGRNFMNPALAGRAFLFFAYPAQIS
                     GDTVWTAADGFSGATALSQWSQGGQGALQHTVTGAPITWMDAFVGNLPGSMGEVSTLA
                     ILIGGAVIVFTRIAAWRIIAGVMIGMIATSTLFNLIGSETNPMFSMPWHWHFVLGGFA
                     LGMVFMATDPVSASFTNTGKWWYGALIGVMAVLIRTVNPAYPEGMMLAILFANLFAPI
                     FDYIVVQANIKRRRARTNG"
gene            5540. .6274
                     /gene="HI0167"
CDS             5540. .6274
                     /gene="HI0167"
                     /note="similar to PID:663269 PID:663272 PID:893413 percent
                     identity: 54.36; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating"
                     /protein_id="AAC21838.1"
                     /db_xref="GI:1573124"
                     /translation="MAKFNKDSVGGTILVVLLLSLVCSIIVAGSAVMLKPAQEEQKLL
                     DKQKNILNVAGLLQENTNVKETYAKFIEPRFVDLATGEYTQQADDSQQAIPADADKAR
                     IRSRSKTTEVYLVKDEQGQTQQVILPIYGTGLWSVMYGLVSVQPDGNTINGITYYQHG
                     ETPGLGGEIENPNWASLFKGKKLFDEQHQPAIRIVKGQAPQDEHSIDGLSGATLTGNG
                     VQGTFNYWFSKDGFGPYLEKLHSGAN"
gene            6274. .6900
                     /gene="HI0168"
CDS             6274. .6900
                     /gene="HI0168"
                     /note="similar to PID:663269 PID:663273 PID:893414 percent
                     identity: 74.88; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating"
                     /protein_id="AAC21839.1"
                     /db_xref="GI:1573125"
                     /translation="MSGKTSYKDLLLAPIAKNNPIALQILGICSALAVTTKLETAFVM
                     AIAVTLVTGLSNLFVSLIRNYIPNSIRIIVQLAIIASLVIVVDQILKAYAYGLSKQLS
                     VFVGLIITNCIVMGRAEAFAMKSPPVESFVDGIGNGLGYGSMLIIVAFFRELIGSGKL
                     FGMTIFETIQNGGWYQANGLFLLAPSAFFIIGFVIWGLRTWKPEQQEK"
gene            6904. .7500
                     /gene="HI0170"
```

```
    CDS             6904. .7500
                    /gene="HI0170"
                    /note="similar to PID:893416 PID:893415 percent identity:
                    84.34; identified by sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="NADH:ubiquinone oxidoreductase, Na
                    translocating"
                    /protein_id="AAC21840.1"
                    /db_xref="GI:1573126"
                    /translation="MEHYISLFVKAVFIENMALSFFLGMCTFLAVSKKVSPAFGLGIA
                    VTFVLGIAVPVNQLIYANVLKENALIEGVDLSFLNFITFIGVIAGLVQILEMVLDKFM
                    PSLYNALGIFLPLIAVNCAIFGGVSFMVQRDYNFPESIVYGFGSGLGWMLAIVALAGL
                    TEKMKYADIPAGLKGLGITFISVGLMALGFMSFSGIQL"

  Query Match             39.4%;  Score 530;  DB 1;  Length 13299;
  Best Local Similarity   62.7%;  Pred. No. 1.7e-118;
  Matches  844;  Conservative    0;  Mismatches  497;  Indels    6;  Gaps    2;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| ||| || |||||| |  | || || || ||||| | ||| |  |||||.||
Db   2966 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 3025

Qy     61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            ||  |||   ||   | |   |||| ||| | || || || || || |||| || |||||
Db   3026 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 3085

Qy    121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          || |  ||||| ||    ||||||||||  || |||||||| ||||| || |||||||||
Db   3086 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTGAAGAC 3145

Qy    181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
          |||||  |||| || ||  | ||||| || || |      || |  ||| | || ||| ||
Db   3146 AAAAAAAATCCTGGTGTAATTTTTACAGCCCCTGCAAGCGGTACCATCACTGCAATCAAT 3205

Qy    241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
          || |||||||| || ||| | || || || || |||   || |||||  | ||  |||||
Db   3206 CGTGGCGAAAAACGTGTATTACAATCTGTTGTTATTAATGTAGAAGGTGATGAAAAAATC 3265

Qy    301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
            || | |   ||     | |||   |||   |   |       ||  || |
Db   3266 ACTTTTGCAAAATATAGCACAGAACAATTGAATACACTTTCTTCTGAACAAGTAAAACAA 3325

Qy    361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
          ||  ||||  |||| || || ||||| ||  | ||||  ||||| || || ||| |.||
Db   3326 AACTTGATAGAATCTGGCTTATGGACGGCATTACGTACTCGTCCATTTAGTAAAGTTCCA 3385

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
            || | ||    ||  | |   | || || || ||||||||||| ||||||||  |.||
Db   3386 TCCATTGAAAGTGAAGCATCTTCTATTTTTGTAAATGCGATGGATACCAATCCATTAGCA 3445

Qy    481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
          || || ||    || ||  | ||||||      | || ||||     || |    |||
Db   3446 GCCGATCCCTCTGTCGTATTAAAAGAATATTCGCAAGACTTCACTAACGGTTTAACCGTA 3505

Qy    541 ttgagccgtttgaccgagcgtaa---aatccatgtgtgtaaggcagctggcgcagacgtg 597
          || || ||| |   |      ||   | | ||  | ||||| ||||  |   |  |  |
Db   3506 TTAAGTCGTCTATTCCCTTCAAAACCATTACACTTATGTAAAGCAGGCGATTCTAATATT 3565

Qy    598 ccgtctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggt 657
          ||  |||   ||  ||  ||   | | |  ||||| ||    ||    ||||| || |||
Db   3566 CCAACTGCTGATCTTGAAAATTTACAGATTCATGATTTTACTGGTGTTCATCCAGCTGGT 3625

Qy    658 ttgagtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggacc 717
            |   || ||||| ||||| || ||||| || ||||||    | ||||| || |||
Db   3626 CTTGTAGGGACGCATATTCACTTTATTGATCCTGTCGGTATTCAAAAAACTGTATGGCAT 3685

Qy    718 atcaattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacacc 777
          ||||||||||||||||| || ||  | ||    || ||| ||||||||   ||  |  |
Db   3686 ATCAATTATCAAGATGTGATCGCTGTAGGTAAATTATTTACAACAGGCGAACTTTATTCA 3745

Qy    778 gagcgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgtt 837
          || || || ||| || | |  ||| ||||||| ||  |||| ||  |  | |||||

Db   3746 GAACGAGTAATTTCTCTTGCAGGTCCTCAAGTGAAAGAACCTCGTTTAGTACGTACAACA 3805

Qy     838 ttgggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtg 897
           | ||||| ||  | || ||| | ||       |||||   ||  |   | || || ||
Db   3806 ATCGGTGCTAATCTTTCTCAATTAACCCAAAATGAATTAAGTGCAGGTAAAAATCGTGTA 3865

Qy     898 atttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgc 957
           ||||| || || ||  |     ||     ||  || |||   | || |||||||| || |||
Db   3866 ATTTCTGGCTCAGTTCTTTGTGGTCAAATAGCAAAAGATTCACATGATTATTTAGGTCGC 3925

Qy     958 taccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggtt 1017
           ||       ||  | || || ||||  |||||       |||||  | || || ||| ||
Db   3926 TATGCTTTGCAAGTATCAGTAATCGCTGAAGGTAATGAAAAAGAATTTTTTGGTTGGATT 3985

Qy    1018 gcgccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaac 1077
             ||| ||  |  | ||||| ||  | || |||||     | ||||| |||   |  ||
Db   3986 ATGCCACAAGCCAATAAATATTCTGTGACTCGTACCGTATTAGGCCACTTC---AGTAAA 4042

Qy    1078 aaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggt 1137
           ||| | || || || || ||  | |  || |||||||| || || |||||||| || |||
Db   4043 AAATTGTTTAATTTTACTACTTCAGAAAATGGTGGCGAACGTGCAATGGTGCCAATCGGT 4102

Qy    1138 acttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatc 1197
           |  || ||||||||||||||  | |||||  | |||||  |  | || || |||||||||
Db   4103 AGCTATGAGCGCGTAATGCCATTGGACATTTTACCTACATTATTATTACGTGATTTAATC 4162

Qy    1198 gtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctc 1257
           || |||||||||||  | || ||||  || |||||  ||||||| ||||||||||||| |
Db   4163 GTGGGCGATACCGATGGTGCTCAAGAGTTAGGTTGTCTGGAATTAGACGAAGAAGACTTG 4222

Qy    1258 gctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtg 1317
           || || ||    ||||| || || |||||||| ||||| ||  |  |  | |||||| | ||
Db   4223 GCGTTATGTTCTTTCGTTTGTCCAGGCAAATATGAATACGGTTCAATCTTGCGTCAAGTC 4282

Qy    1318 ctggaaaccnttgagaaggaaggctga 1344
           | || |   ||||||||||||| | |
Db   4283 TTAGATAAGATTGAGAAGGAAGGTTAA 4309


RESULT  10
AF165980
LOCUS       AF165980               6048 bp    DNA     linear   BCT 13-NOV-2001
DEFINITION  Vibrio harveyi Na+-translocating NADH-quinone oxidoreductase
            complex operon, complete sequence.
ACCESSION   AF165980
VERSION     AF165980.1  GI:6531420
KEYWORDS    .
SOURCE      Vibrio harveyi.
  ORGANISM  Vibrio harveyi
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 6048)
  AUTHORS   Zhou,W., Bertsova,Y.V., Feng,B., Tsatsos,P., Verkhovskaya,M.L.,
            Gennis,R.B., Bogachev,A.V. and Barquera,B.
  TITLE     Sequencing and preliminary characterization of the
            Na+-translocating NADH: ubiquinone oxidoreductase from Vibrio
            harveyi
  JOURNAL   Biochemistry 38 (49), 16246-16252 (1999)
  MEDLINE   20056044
   PUBMED   10587447
REFERENCE   2  (bases 1 to 6048)
  AUTHORS   Zhou,W., Feng,B., Gennis,R.B. and Barquera,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1999) Biochemistry, University of Illinois, 600
            South Mathews Avenue, Urbana, IL 61801, USA
FEATURES             Location/Qualifiers
     source          1. .6048
                     /organism="Vibrio harveyi"
                     /strain="BB120"
                     /db_xref="taxon:669"
     gene            169. .1509
                     /gene="nqrA"
     gene            169. .6022
                     /gene="Na+-translocating NADH-quinone oxidoreductase
```

```
                complex operon"
     CDS        169. .1509
                /gene="nqrA"
                /codon_start=1
                /transl_table=11
                /product="NqrA"
                /protein_id="AAF15411.1"
                /db_xref="GI:6531421"
                /translation="MITIKKGLALPIAGAPSQVINDGKTIKKVALLGEEYVGMRPTMH
                VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEINRGAKRVLQSVVIEVAGEEQVTF
                DKFEAAQLAGLDREVIKTQLVESGLWTALRTRPFSKVPAIESATKAIFVTAMDTNPLA
                AQPELIITEQQEAFVAGLDILSALTEGKVYVCKSGTSLPSSSQSNVEEHVFDGPHPAG
                LAGTHMHFLYPVNAENVAWSINYQDVIAFGQLFLTGELYTDRVVSLAGPVVNNPRLLR
                TVVGASLEDLTDSELMPGEVRVISGSVLSGTQASGPHAYLGRYHQQVSVLREGRDKEL
                FGWATPGKNKFSITKSFLGHIFKGQLFNMTTTTNGSDRAMVPIGNYERVMPLDMEPTL
                LLRDLCAGDTDSAQTLGALELDEEDLALCTFVCPGKYEYGQLLRECLDTIVKEG"
     gene       1513. .2754
                /gene="nqrB"
     CDS        1513. .2754
                /gene="nqrB"
                /codon_start=1
                /transl_table=11
                /product="NqrB"
                /protein_id="AAF15412.1"
                /db_xref="GI:6531422"
                /translation="MLKKFIEDIEHHFEPGGKHEKWFALYEAAATLFYTPGLVTKRSS
                HVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGDQLATVIAGNWHYW
                LTEMLGGTIGAEAGVGSMMLLGATYFLPIYATVFIVGGFWEVLFCMVRKHEVNEGFFV
                TSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPALAGRAFLFFAYPAQ
                ISGDVVWTAADGFSGATALSQWAHRGSGALINNITGAPITWMDAFIGNIPGSIGEVST
                LALMIGAAMIVYMRIASWRIIAGVMIGMIAVSTLFNVIGSDTNPMFNMPWHWHLVLGG
                FAFGMFFMATDPVSASFTNSGKWWYGILIGAMCVMIRVVNPAYPEGMMLAILFANLFA
                PLFDHVVIEKNIKRRLARYGK"
     gene       2744. .3529
                /gene="nqrC"
     CDS        2744. .3529
                /gene="nqrC"
                /codon_start=1
                /transl_table=11
                /product="NqrC"
                /protein_id="AAF15413.1"
                /db_xref="GI:6531423"
                /translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDQQKANAVLD
                KQSKIVEVAGIDAEGKKVPELFAEYIEPRLVDFKTGDFVEKAEDGSTAANYDQRKAAK
                DPAESIKLTADEDKAKILRRANTGIVYLVKNGDDISKVIIPVHGNGLWSMMYAFVAVE
                TDGNTVSGITYYEQGETPGLGGEVENPVWRAQFVGKKLFDENHKPAIKIVKGGAPEGS
                EHGVDGLSGATLTGNGVQGTFDFWLGDMGFGPFLAKVRDGGLN"
     gene       3529. .4161
                /gene="nqrD"
     CDS        3529. .4161
                /gene="nqrD"
                /codon_start=1
                /transl_table=11
                /product="NqrD"
                /protein_id="AAF15414.1"
                /db_xref="GI:6531424"
                /translation="MSSAQNIKKSIMAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
                VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                FGMEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
     gene       4169. .4765
                /gene="nqrE"
     CDS        4169. .4765
                /gene="nqrE"
                /codon_start=1
                /transl_table=11
                /product="NqrE"
                /protein_id="AAF15415.1"
                /db_xref="GI:6531425"
                /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGVA
                VVVVLTIAVPVNNLVYNLVLKENALVEGVDLSFLNFITFIGVIAALVQILEMVLDRFF
                PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYNFAESIVYGFGSGVGWMLAIVALAGI
                REKMKYSDVPPGLRGLGITFITVGLMALGFMSFSGVQL"
     gene       4799. .6022
                /gene="nqrF"
     CDS        4799. .6022
                /gene="nqrF"
                /codon_start=1
                /transl_table=11
                /product="NqrF"
                /protein_id="AAF15416.1"
                /db_xref="GI:6531426"
                /translation="MDIILGVVMFTLIVLALVLVILFAKSKLVPTGDITISVNDDPSL
                AIVTQPGGKLLSALAGAGVFVSSACGGGGSCGQCRRKIKSGGGDILPTELDHITKGEA
                REGERLACQVAMKTDMDIELPEEIFGVKKWECTVISNDNKATFIKELKLQIPDGESVP
                FRAGGYIQIEAPAHHVKYADYDIPEEYREDWEKFNLFRYESKVNEETIRAYSMANYPE
                EHGIIMLNVRIATPPPNNPDVAPGIMSSFIWSLKEGDKCTISGPFGEFFAKDTDAEMV
                FVGGGAGMAPMRSHIFDQLKRLHSKRKMSFWYGARSKREMFYVEDFDGLAADNDNFVW
                HCALSDPLPEDNWDGYTGFIHNVLYENYLRDHEAPEDCEYYMCGPPMMNAAVIGMLKD
                LGVEDENILLDDFGG"
BASE COUNT     1571 a   1351 c   1455 g   1671 t
ORIGIN
```

```
  Query Match             37.7%;  Score 506.2;  DB 1;  Length 6048;
  Best Local Similarity   61.1%;  Pred. No. 1.1e-112;
  Matches  821;  Conservative    0;  Mismatches  520;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| ||| || || ||  |    || || ||||| ||    ||    || || |||
Db     169 ATGATTACAATAAAGAAGGGCTTGGCTCTTCCTATCGCAGGAGCTCCATCCCAGGTGATT 228

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            |||| ||      ||| |   ||||||| ||||||||||||||| || |  || |||||
Db     229 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 288

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||    ||| | |||   |  |||||||  || |||||||  ||||| || |||||||||
Db     289 CCTACCATGCATGTCCGCGTTGGCGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAC 348

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
           |||||| | || |||||   ||| || || || |   | || ||| |   ||  ||  |
Db     349 AAAAAGAACCCTGGCGTTAAGTTCACTGCTCCAGCAGCCGGTAAAGTGATCGAAATTAAC 408

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
           || ||||  ||||| |||||||| || || |||||||  || | |||| | ||  |  |
Db     409 CGTGGCGCTAAGCGTGTACTTCAATCAGTTGTGATTGAAGTAGCAGGCGAAGAGCAGGTT 468

Qy     301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
           |||  ||    | ||   | |      | |||   ||    ||||  |   |
Db     469 ACGTTTGATAAGTTCGAAGCAGCTCAACTAGCAGGTCTAGATCGCGAAGTGATCAAGACT 528

Qy     361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
           |  || |  |||| ||  | ||||| ||  | ||||  ||||||||||||||  | ||
Db     529 CAATTGGTTGAATCTGGCCTTTGGACCGCTTTACGTACTCGTCCGTTCAGCAAGGTTCCA 588

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
           ||  |||| | |    |    || || ||||| | |||||||||| || ||||| || ||
Db     589 GCAATCGAGTCTGCTACTAAGGCTATTTTCGTAACTGCGATGGATACTAATCCACTAGCA 648

Qy     481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
           ||| | |||| | |  | |||| |||      || |  |||     | || |     |
Db     649 GCACAGCCTGAGTTGATCATCACTGAACAACAAGAAGCATTCGTTGCAGGTCTAGACATT 708

Qy     541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
           |        |||| ||  | ||  |  | || ||||||    ||||| |   | ||||
Db     709 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAAG---TCTGGCACTAGCTTGCCA 765

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
                 |   | ||| | |||  ||| |  |||| || || || || || ||| |
Db     766 AGCTCTTCTCAGTCAAACGTAGAAGAACACGTCTTCGATGGTCCTCACCCAGCAGGTCTT 825

Qy     661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
           ||||||  ||||| |||||| |  | || ||   |||| | ||     |  |||| ||||
Db     826 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCGTGGAGCATC 885
```

```
Qy     721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
           || || ||||| || || ||  |||| |   | ||   ||| ||    ||  |||| ||
Db     886 AACTACCAAGACGTTATCGCGTTCGGTCAGCTTTTCCTAACGGGTGAGCTTTACACTGAT 945

Qy     781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
           ||  ||   |  || | | |||| ||   || ||||| ||||||||  | |||| ||| |
Db     946 CGTGTTGTGTCTCTAGCTGGTCCTGTTGTGAACAACCCACGCCTTCTTCGTACAGTTGTA 1005

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
           |||||     |     |  | ||||   |||| ||| |    |  ||    || ||||||
Db    1006 GGTGCTTCTCTTGAAGACCTAACTGACAGCGAGTTGATGCCTGGTGAAGTTCGTGTGATT 1065

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
           || || |||||| |    ||  |     |    ||  | || |  ||  | || || |||
Db    1066 TCAGGCTCGGTACTTTCTGGTACTCAAGCGTCTGGTCCACATGCCTACCTAGGTCGTTAC 1125

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||| | ||  |||| ||  |    |||||  ||   |||||| ||||| ||||||| | |
Db    1126 CACCAACAAGTTTCTGTGTTACGCGAAGGTCGTGACAAAGAACTGTTTGGCTGGGCTACA 1185

Qy    1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
           ||       |||| | ||| ||||||    |   ||| || ||  || | ||     |
Db    1186 CCTGGTAAAAACAAGTTCTCAATCACGAAATCTTTCCTTGGTCACATCTTCAAGGGTCAG 1245

Qy    1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
           | |||||  | || ||| |   |||||| | ||||| || ||||| || || || |
Db    1246 TTGTTCAACATGACAACAACGACCAACGGTAGTGACCGTGCAATGGTTCCAATCGGCAAC 1305

Qy    1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
           |||||||||||||||||||| || ||    || ||  |||  | || ||| ||   |
Db    1306 TACGAGCGCGTAATGCCGCTTGATATGGAACCGACTTTGCTACTTCGTGATCTATGTGCA 1365

Qy    1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
           |||||||| ||||| || ||| || |||||    | ||| |||| |||||||| || ||
Db    1366 GGCGATACAGACAGTGCTCAAACACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 1425

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
           ||||| | ||| || || || || ||||||||| || || | |||  | ||| |    ||
Db    1426 TTGTGTACCTTTGTATGTCCTGGTAAATACGAGTACGGTCAGCTACTTCGTGAATGCCTA 1485

Qy    1321 gaaaccnttgagaaggaaggctga 1344
           || ||| | | ||||||||| | |
Db    1486 GATACCATCGTGAAGGAAGGGTAA 1509


RESULT  11
AF117331
LOCUS       AF117331                5916 bp    DNA     linear   BCT 13-MAY-1999
DEFINITION  Vibrio cholerae N16961 Na+-translocating NADH-ubiquinone
            oxidoreductase enzyme complex, complete sequence.
ACCESSION   AF117331
VERSION     AF117331.1  GI:4808841
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 5916)
  AUTHORS   Hase,C.C. and Mekalanos,J.J.
  TITLE     Effects of changes in membrane sodium flux on virulence gene
            expression in Vibrio cholerae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3183-3187 (1999)
  MEDLINE   99179036
REFERENCE   2  (bases 1 to 5916)
  AUTHORS   Hase,C.C. and Mekalanos,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1998) Microbiology and Molecular Genetics,
            Harvard Medical School, 200 Longwood Ave., Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .5916
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /db_xref="taxon:666"
                     /note="biotype: El Tor"
     gene            34. .1374
                     /gene="nqrA"
     CDS             34. .1374
                     /gene="nqrA"
                     /note="alpha subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrA"
                     /protein_id="AAD29962.1"
                     /db_xref="GI:4808842"
                     /translation="MITIKKGLDLPIAGTPSQVISDGKAIKKVALLGEEYVGMRPTMH
                     VRVGDEVKKAQILFEDKKNPGVKFTSPVSGKVVEINRGAKRVLQSVVIEVAGDDQVTF
                     DKFEANQLASLNRDAIKTQLVESGLWTAFRTRPFSKVPAIDSTSEAIFVTAMDTNPLA
                     AEPTVVINEQSEAFVAGLDVLSALTTGKVYVCKKGTSLPRSQQPNVEEHVFDGPHPAG
                     LAGTHMHFLYPVSADHVAWSINYQDVIAVGQLFLTGELYTQRVVSLAGPVVNKPRLVR
                     TVMGASLEQLVDSEIMPGEVRIISGSVLSGTKATGPHAYLGRYHLQVSVLREGRDKEL
                     FGWAMPGKNKFSVTRSFLGHLFKGQVYNMTTTTNGSDRSMVPIGNYEKVMPLDMEPTL
                     LLRDLCAGDSDSAVRLGALELDEEDLALCTFVCPGKYEYGQLLRECLDKIEKEG"
     gene            1380. .2627
                     /gene="nqrB"
     CDS             1380. .2627
                     /gene="nqrB"
                     /note="a subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrB"
                     /protein_id="AAD29963.1"
                     /db_xref="GI:4808843"
                     /translation="MGLKKFLEDIEHHFEPGGKHEKWFALYEAAATLFYTPGLVTKRS
                     SHVRDSVDLKRIMIMVWLAVFPAMFWGMYNAGGQAIAALNHLYSGDQLAAIVAGNWHY
                     WLTEMLGGTMSSDAGWGSKMLLGATYFLPIYATVFIVGGFWEVLFCMVRKHEVNEGFF
                     VTSILFALIVPPTLPLWQAALGITFGVVVAKEVFGGTGRNFLNPALAGRAFLFFAYPA
                     QISGDLVWTAADGYSGATALSQWAQGGAGALINNATGQTITWMDAFIGNIPGSIGEVS
                     TLALMIGAAFIVYMGIASWRIIGGVMIGMILLSTLFNVIGSDTNAMFNMPWHWHLVLG
                     GFAFGMFFMATDPVSASFTNSGKWAYGILIGVMCVLIRVVNPAYPEGMMLAILFANLF
                     APLFDHVVVERNIKRRLARYGKQ"
     gene            2614. .3387
                     /gene="nqrC"
     CDS             2614. .3387
                     /gene="nqrC"
                     /note="gamma subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrC"
                     /protein_id="AAD29964.1"
                     /db_xref="GI:4808844"
                     /translation="MASNNDSIKKTLFVVIALSLVCSIIVSAAAVGLRDKQKENAALD
                     KQSKILQVAGIEAKGSKQIVELFNKSIEPRLVDFNTGDFVEGDAANYDQRKAAKEASE
                     SIKLTAEQDKAKIQRRANVGVVYLVKDGDKTSKVILPVHGNGLWSMMYAFVAVETDGN
                     TVSGLTYYEQGETPGLGGEVENPAWRAQWVGKKLFDENHKPAIKIVKGGAPQGSEHGV
                     DGLSGATLTSNGVQNTFDFWLGDMGFGPFLTKVRDGGLN"
     gene            3387. .4019
                     /gene="nqrD"
     CDS             3387. .4019
                     /gene="nqrD"
                     /note="b subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrD"
                     /protein_id="AAD29965.1"
                     /db_xref="GI:4808845"
                     /translation="MSSAKELKKSVLAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVMFVTALSNFFVSLIRNHIPNSVRIIVQMAIIASLVIVVDQILKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSEPIPSFIDGIGNGLGYGFVLMTVGFFRELLGSGKL
                     FGLEVLPLISNGGWYQPNGLMLLAPSAFFLIGFMIWAIRTFKPEQVEAKE"
     gene            4027. .4623
                     /gene="nqrE"
     CDS             4027. .4623
                     /gene="nqrE"
                     /note="c subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrE"
```

```
                     /protein_id="AAD29966.1"
                     /db_xref="GI:4808846"
                     /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGIA
                     VIVVLTISVPVNNLVYNLVLKPDALVEGVDLSFLNFITFIGVIAALVQILEMILDRFF
                     PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYSFAESVVYGFGSGVGWMLAIVALAGI
                     REKMKYSDVPPGLRGLGITFITAGLMALGFMSFSGVQL"
     gene            4657. .5883
                     /gene="nqrF"
     CDS             4657. .5883
                     /gene="nqrF"
                     /note="beta subunit of NQR enzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrF"
                     /protein_id="AAD29967.1"
                     /db_xref="GI:4808847"
                     /translation="MSTIIFGVVMFTLIILALVLVILFAKSKLVPTGDITISINGDPE
                     KAIVTQPGGKLLTALAGAGVFVSSACGGGGSCGQCRVKIKSGGGDILPTELDHISKGE
                     AREGERLACQVAVKADMDLELPEEIFGVKKWECTVISNDNKATFIKELKLAIPDGESV
                     PFRAGGYIQIEAPAHHVKYADFDVPEKYRGDWDKFNLFRYESKVDEPIIRAYSMANYP
                     EEFGIIMLNVRIATPPPNNPNVPPGQMSSYIWSLKAGDKCTISGPFGEFFAKDTDAEM
                     VFIGGGAGMAPMRSHIFDQLKRLKSKRKMSYWYGARSKREMFYVEDFDGLAAENDNFV
                     WHCALSDPQPEDNWTGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPMMNAAVINMLK
                     NLGVEEENILLDDFGG"
BASE COUNT     1441 a   1276 c   1550 g   1649 t
ORIGIN
```

```
  Query Match             37.4%;  Score 503;  DB 1;  Length 5916;
  Best Local Similarity   60.9%;  Pred. No. 6.6e-112;
  Matches  819;  Conservative    0;  Mismatches  522;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
Db      34 ATGATTACAATAAAAAAGGGATTGGACCTTCCTATCGCAGGAACTCCTTCCCAGGTGATT 93

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
Db      94 AGTGATGGTAAAGCCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 153

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
Db     154 CCTACTATGCATGTCCGCGTAGGTGATGAAGTAAAAAAAGCTCAGATTCTTTTTGAAGAT 213

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
Db     214 AAGAAGAATCCGGGTGTGAAATTTACTTCACCAGTCAGCGGTAAAGTGGTTGAAATCAAT 273

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
Db     274 CGTGGTGCAAAGCGCGTTCTTCAGTCTGTAGTGATTGAAGTGGCGGGTGATGACCAAGTG 333

Qy     301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
Db     334 ACCTTCGACAAGTTTGAAGCCAATCAATTAGCCAGTTTGAATCGCGATGCGATCAAAACT 393

Qy     361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
Db     394 CAGCTGGTTGAGTCTGGTCTTTGGACTGCTTTCCGTACTCGTCCGTTCAGCAAGGTTCCA 453

Qy     421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
Db     454 GCCATTGATTCCACTTCTGAAGCCATTTTTGTCACTGCTATGGATACCAATCCGTTGGCA 513

Qy     481 gcagaccctgtggttgtgatcaaagaagccgngcgangatttcagacgangtntgctggta 540
Db     514 GCTGAGCCGACCGTTGTGATCAATGAACAATCTGAAGCCTTCGTTGCTGGTCTAGATGTG 573

Qy     541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
Db     574 CTTTCCGCACTGACAACTGGCAAAGTGTACGTTTGTAA---AAAAGGCACCAGCTTGCCA 630

Qy     601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
Db     631 CGTTCTCAACAGCCGAATGTTGAAGAACATGTCTTTGACGGCCCACATCCTGCAGGCTTA 690

Qy     661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
Db     691 GCTGGCACTCACATGCATTTCCTCTATCCAGTAAGTGCAGACCACGTGGCGTGGAGCATC 750

Qy     721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
Db     751 AACTATCAGGACGTGATCGCAGTAGGTCAACTGTTCCTGACGGGCGAGCTATACACCCAG 810

Qy     781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
Db     811 CGTGTGGTTTCTCTTGCCGGCCCGGTGGTGAACAAACCTCGTTTGGTTCGCACTGTCATG 870

Qy     841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
Db     871 GGCGCATCACTAGAGCAACTGGTGGACAGCGAAATCATGCCTGGCGAAGTACGCATCATC 930

Qy     901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
Db     931 TCTGGCTCTGTGCTGTCTGGTACGAAAGCCACTGGACCACACGCTTATCTTGGTCGTTAC 990

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db     991 CACCTACAGGTCTCTGTTCTTCGTGAAGGACGTGATAAAGAACTGTTTGGTTGGGCGATG 1050

Qy    1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
Db    1051 CCGGGTAAGAACAAGTTCTCTGTGACTCGCTCTTTCCTTGGTCACCTATTCAAGGGCCAA 1110

Qy    1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
Db    1111 GTGTACAACATGACCACAACCACTAATGGTAGTGATCGATCTATGGTTCCAATCGGTAAC 1170

Qy    1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
Db    1171 TATGAGAAGGTCATGCCACTAGATATGGAGCCAACCTTGTTGCTTCGTGATCTGTGTGCT 1230

Qy    1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
Db    1231 GGTGACAGTGACAGCGCGGTTCGCTTAGGTGCTTTAGAGCTGGATGAAGAAGATCTAGCA 1290

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
Db    1291 CTGTGTACGTTCGTTTGCCCAGGCAAATACGAGTACGGTCAGCTCCTGCGTGAGTGCTTA 1350

Qy    1321 gaaaccnttgagaaggaaggctga 1344
Db    1351 GACAAGATCGAGAAAGAAGGGTAA 1374
```

```
RESULT  12
AE004300/c
LOCUS       AE004300               13382 bp    DNA     linear   BCT 31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 208 of 251 of the complete
            chromosome.
ACCESSION   AE004300 AE003852
VERSION     AE004300.1  GI:9656850
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 13382)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
```

```
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 13382)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .13382
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(154. .1236)
                     /gene="VC2287"
     CDS             complement(154. .1236)
                     /gene="VC2287"
                     /note="similar to GB:U00096 SP:Q47155 PID:1552799
                     PID:1786425; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-damage-inducible protein P"
                     /protein_id="AAF95431.1"
                     /db_xref="GI:9656851"
                     /translation="MQDRIRKIIHVDMDCFFAAVEMRDNPAYREIALAVGGHEKQRGV
                     ISTCNYQARKFGVRSAMPTAQALKLCPQLHVVPGRMSVYKSVSQQIQTIFQRYTSLIE
                     PLSLDEAYLDVSESTAYQGSATLIAQAIRRDIWQELNLTASAGVAPIKFLAKVASDLN
                     KPDGLYVVTPDKVQEMVDSLPLEKIPGVGKVALEKLHQAGLYVGADVRRADYRKLLHQ
                     FGRLGASLWKKSHGIDEREVVTERERKSVGVEYTFSQNISTFQECWQVIEQKLYPELD
                     ARLSRAHPQRGIIKQGIKVKFADFQQTTIEHVHPALELDYFHELLEQVLTRQQGREIR
                     LLGLSVMLKPELQMKQLSMFPSDGWQ"
     gene            complement(1345. .1578)
                     /gene="VC2288"
     CDS             complement(1345. .1578)
                     /gene="VC2288"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF95432.1"
                     /db_xref="GI:9656852"
                     /translation="MNTFLITFGFFLAVIAAMAVGYIFQRKVVKGSCGGLGAVGIEKV
                     CNCPEPCDARKKREAKAAARAERLAAWEKDRIA"
     gene            complement(1595. .2698)
                     /gene="VC2289"
     CDS             complement(1595. .2698)
                     /gene="VC2289"
                     /note="similar to GB:L42023 SP:P44550 PID:1003244
                     PID:1222087 PID:1204428; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="thiamin biosynthesis lipoprotein ApbE"
                     /protein_id="AAF95433.1"
                     /db_xref="GI:9656853"
                     /translation="MFFTNWILKVSMVIWFRVSVFPISIFNYERSKQVRNWLVALASL
                     LLLAGCEKPAEQVHLSGPTMGTTYNIKYIQQPGIADSKTLQTEIDRLLEEVNDQMSTY
                     RKDSELSRFNQHTSSEPFAVSTQTLTVVKEAIRLNGLTEGALDVTVGPLVNLWGFGPE
                     ARPDVVPTDEELNARRAITGIEHLTIEGNTLSKDIPELYVDLSTIAKGWGVDVVADYL
                     QSQGIENYMVEIGGEIRLKGLNRDGVPWRIAIEKPSVDQRSVQEIIEPGDYAIATSGD
                     YRNYFEQDGVRYSHIIDPTTGRPINNRVVSVTVLDKSCMTADGLATGLMVMGEERGMA
                     VAEANQIPVLMIVKTDDGFKEYASSSFKPFLSK"
     gene            complement(2745. .3971)
                     /gene="VC2290"
     CDS             complement(2745. .3971)
                     /gene="VC2290"
                     /note="similar to PID:893416; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, beta subunit"
                     /protein_id="AAF95434.1"
                     /db_xref="GI:9656854"
                     /translation="MSTIIFGVVMFTLIILALVLVILFAKSKLVPTGDITISINGDPE
                     KAIVTQPGGKLLTALAGAGVFVSSACGGGGSCGQCRVKIKSGGGDILPTELDHISKGE
                     AREGERLACQVAVKADMDLELPEEIFGVKKWECTVISNDNKATFIKELKLAIPDGESV
                     PFRAGGYIQIEAPAHHVKYADFDVPEKYRGDWDKFNLFRYESKVDEPIIRAYSMANYP
                     EEFGIIMLNVRIATPPPNNPNVPPGQMSSYIWSLKAGDKCTISGPFGEFFAKDTDAEM
                     VFIGGGAGMAPMRSHIFDQLKRLKSKRKMSYWYGARSKREMFYVEDFDGLAAENDNFV
                     WHCALSDPQPEDNWTGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPMMNAAVINMLK
                     NLGVEEENILLDDFGG"
     gene            complement(4005. .4601)
                     /gene="VC2291"
     CDS             complement(4005. .4601)
                     /gene="VC2291"
                     /note="similar to PID:893416 PID:893415; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, hydrophobic membrane protein NqrE"
                     /protein_id="AAF95435.1"
                     /db_xref="GI:9656855"
                     /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGIA
                     VIVVLTISVPVNNLVYNLVLKPDALVEGVDLSFLNFITFIGVIAALVQILEMILDRFF
                     PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYSFAESVVYGFGSGVGWMLAIVALAGI
                     REKMKYSDVPPGLRGLGITFITAGLMALGFMSFSGVQL"
     gene            complement(4609. .5241)
                     /gene="VC2292"
     CDS             complement(4609. .5241)
                     /gene="VC2292"
                     /note="similar to PID:663269 PID:663273 PID:893414;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, hydrophobic membrane protein NqrD"
                     /protein_id="AAF95436.1"
                     /db_xref="GI:9656856"
                     /translation="MSSAKELKKSVLAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVMFVTALSNFFVSLIRNHIPNSVRIIVQMAIIASLVIVVDQILKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSEPIPSFIDGIGNGLGYGFVLMTVGFFRELLGSGKL
                     FGLEVLPLISNGGWYQPNGLMLLAPSAFFLIGFMIWAIRTFKPEQVEAKE"
     gene            complement(5241. .6014)
                     /gene="VC2293"
     CDS             complement(5241. .6014)
                     /gene="VC2293"
                     /note="similar to PID:663269 PID:663272 PID:893413;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, gamma subunit"
                     /protein_id="AAF95437.1"
                     /db_xref="GI:9656857"
                     /translation="MASNNDSIKKTLFVVIALSLVCSIIVSAAAVGLRDKQKENAALD
                     KQSKILQVAGIEAKGSKQIVELFNKSIEPRLVDFNTGDFVEGDAANYDQRKAAKEASE
                     SIKLTAEQDKAKIQRRANVGVVYLVKDGDKTSKVILPVHGNGLWSMMYAFVAVETDGN
                     TVSGLTYYEQGETPGLGGEVENPAWRAQWVGKKLFDENHKPAIKIVKGGAPQGSEHGV
                     DGLSGATLTSNGVQNTFDFWLGDMGFGPFLTKVRDGGLN"
     gene            complement(6001. .7284)
                     /gene="VC2294"
     CDS             complement(6001. .7284)
                     /gene="VC2294"
                     /note="similar to PID:663269 PID:663271; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, hydrophobic membrane protein NqrB"
                     /protein_id="AAF95438.1"
                     /db_xref="GI:9656858"
                     /translation="MLRQDRERRVISMGLKKFLEDIEHHFEPGGKHEKWFALYEAAAT
                     LFYTPGLVTKRSSHVRDSVDLKRIMIMVWLAVFPAMFWGMYNAGGQAIAALNHLYSGD
                     QLAAIVAGNWHYWLTEMLGGTMSSDAGWGSKMLLGATYFLPIYATVFIVGGFWEVLFC
                     MVRKHEVNEGFFVTSILFALIVPPTLPLWQAALGITFGVVVAKEVFGGTGRNFLNPAL
                     AGRAFLFFAYPAQISGDLVWTAADGYSGATALSQWAQGGAGALINNATGQTITWMDAF
                     IGNIPGSIGEVSTLALMIGAAFIVYMGIASWRIIGGVMIGMILLSTLFNVIGSDTNAM
                     FNMPWHWHLVLGGFAFGMFFMATDPVSASFTNSGKWAYGILIGVMCVLIRVVNPAYPE
                     GMMLAILFANLFAPLFDHVVVERNIKRRLARYGKQ"
     gene            complement(7254. .8642)
                     /gene="VC2295"
     CDS             complement(7254. .8642)
                     /gene="VC2295"
                     /note="similar to PID:663269 PID:663270; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating, alpha subunit"
                     /protein_id="AAF95439.1"
                     /db_xref="GI:9656859"
                     /translation="MRNERIFFPIKVVQVRMITIKKGLDLPIAGTPSQVISDGKAIKK
                     VALLGEEYVGMRPTMHVRVGDEVKKAQILFEDKKNPGVKFTSPVSGKVVEINRGAKRV
                     LQSVVIEVAGDDQVTFDKFEANQLASLNRDAIKTQLVESGLWTAFRTRPFSKVPAIDS
                     TSEAIFVTAMDTNPLAAEPTVVINEQSEAFVAGLDVLSALTTGKVYVCKKGTSLPRSQ
                     QPNVEEHVFDGPHPAGLAGTHMHFLYPVSADHVAWSINYQDVIAVGQLFLTGELYTQR
                     VVSLAGPVVNKPRLVRTVMGASLEQLVDSEIMPGEVRIISGSVLSGTKATGPHAYLGR
                     YHLQVSVLREGRDKELFGWAMPGKNKFSVTRSFLGHLFKGQVYNMTTTTNGSDRSMVP
                     IGNYEKVMPLDMEPTLLLRDLCAGDSDSAVRLGALELDEEDLALCTFVCPGKYEYGQL
                     LRECLDKIEKEG"
     gene            complement(9079. .9399)
                     /gene="VC2296"

  Query Match             37.4%;  Score 503;  DB 1;  Length 13382;
  Best Local Similarity   60.9%;  Pred. No. 6.7e-112;
  Matches  819;  Conservative    0;  Mismatches  522;  Indels    3;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| ||| ||||| ||  |  |||| || ||||| || |  ||    || || |||
Db   8594 ATGATTACAATAAAAAAGGGATTGGACCTTCCTATCGCAGGAACTCCTTCCCAGGTGATT 8535

Qy     61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            ||| ||    | ||| |   |||||||| ||||||||||||||| || |  || |||||
Db   8534 AGTGATGGTAAAGCCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 8475

Qy    121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||    ||| | |||   | ||| ||||  || ||||||| ||  | || ||||||||
Db   8474 CCTACTATGCATGTCCGCGTAGGTGATGAAGTAAAAAAAGCTCAGATTCTTTTTGAAGAT 8415

Qy    181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
          || ||| |||||||| |||   |||||  | || ||    || ||| | |  |  ||| ||
Db   8414 AAGAAGAATCCGGGTGTGAAATTTACTTCACCAGTCAGCGGTAAAGTGGTTGAAATCAAT 8355

Qy    241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
          || || | |||||||||| |||||||| || |||||||  || |  ||  | ||| || |
Db   8354 CGTGGTGCAAAGCGCGTTCTTCAGTCTGTAGTGATTGAAGTGGCGGGTGATGACCAAGTG 8295

Qy    301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
             |||||    |  |   || |    || || |  || |   |||| |   |
Db   8294 ACCTTCGACAAGTTTGAAGCCAATCAATTAGCCAGTTTGAATCGCGATGCGATCAAAACT 8235

Qy    361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
            | ||| |  | || ||| | ||||||||  | |||  |||||||||||||  | ||
Db   8234 CAGCTGGTTGAGTCTGGTCTTTGGACTGCTTTCCGTACTCGTCCGTTCAGCAAGGTTCCA 8175

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
          ||| | ||| ||    |    ||||| || |||| ||| ||||| ||||||||| | ||
Db   8174 GCCATTGATTCCACTTCTGAAGCCATTTTTGTCACTGCTATGGATACCAATCCGTTGGCA 8115

Qy    481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
             || || ||    |||||||||| |||      || |  |||       || |    ||
Db   8114 GCTGAGCCGACCGTTGTGATCAATGAACAATCTGAAGCCTTCGTTGCTGGTCTAGATGTG 8055

Qy    541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
            |   |    ||||     | ||| |  | || |||||   |   ||| |   | ||||
Db   8054 CTTTCCGCACTGACAACTGGCAAAGTGTACGTTTGTAA---AAAAGGCACCAGCTTGCCA 7998

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
            |    |     | ||  | |||  |||||  || | |||||| ||||| || || ||
Db   7997 CGTTCTCAACAGCCGAATGTTGAAGAACATGTCTTTGACGGCCCACATCCTGCAGGCTTA 7938

Qy    661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
            |||||| ||||| |||||| |  | || ||  |||||  || |    |  |||| ||||
Db   7937 GCTGGCACTCACATGCATTTCCTCTATCCAGTAAGTGCAGACCACGTGGCGTGGAGCATC 7878

Qy    721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
          || ||||| || || || ||  | || |   ||||    || |||   ||  ||||| ||
Db   7877 AACTATCAGGACGTGATCGCAGTAGGTCAACTGTTCCTGACGGGCGAGCTATACACCCAG 7818

Qy    781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
          || ||| || || | |  ||  |    || |||||||| ||  |  | || || ||  ||
Db   7817 CGTGTGGTTTCTCTTGCCGGCCCGGTGGTGAACAAACCTCGTTTGGTTCGCACTGTCATG 7758

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
          || ||   | ||  |||| |    |  ||||| |  |    |  ||    ||| | ||
Db   7757 GGCGCATCACTAGAGCAACTGGTGGACAGCGAAATCATGCCTGGCGAAGTACGCATCATC 7698

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
          || || || ||  ||   ||  |||   |    ||  | |||| |||| | || |||
Db   7697 TCTGGCTCTGTGCTGTCTGGTACGAAAGCCACTGGACCACACGCTTATCTTGGTCGTTAC 7638

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          |||   ||| | || ||| |    ||||| ||    ||||| ||||| || ||||    |
Db   7637 CACCTACAGGTCTCTGTTCTTCGTGAAGGACGTGATAAAGAACTGTTTGGTTGGGCGATG 7578

Qy   1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
          |||     | |||| | |||  | || ||  |   ||| || ||  |  | ||   | ||
Db   7577 CCGGGTAAGAACAAGTTCTCTGTGACTCGCTCTTTCCTTGGTCACCTATTCAAGGGCCAA 7518

Qy   1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
            | | |||  | || ||| ||   || ||| | || ||  | ||||| || || ||||
Db   7517 GTGTACAACATGACCACAACCACTAATGGTAGTGATCGATCTATGGTTCCAATCGGTAAC 7458

Qy   1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
          || |||   || ||||| ||||| ||   ||| ||| || |  | || ||| |    |
Db   7457 TATGAGAAGGTCATGCCACTAGATATGGAGCCAACCTTGTTGCTTCGTGATCTGTGTGCT 7398

Qy   1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
          || || |  |||||||||      || |||   || ||  |||| |||||||| || ||
Db   7397 GGTGACAGTGACAGCGCGGTTCGCTTAGGTGCTTTAGAGCTGGATGAAGAAGATCTAGCA 7338

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
            |||| |  ||||| ||||| |||||||||||| || || | |||  ||||| ||    |
Db   7337 CTGTGTACGTTCGTTTGCCCAGGCAAATACGAGTACGGTCAGCTCCTGCGTGAGTGCTTA 7278

Qy   1321 gaaaccnttgagaaggaaggctga 1344
          || |   | ||||| ||||| | |
Db   7277 GACAAGATCGAGAAAGAAGGGTAA 7254


RESULT  13
VANQRBOL
LOCUS       VANQRBOL                5078 bp    DNA     linear   BCT 15-JUN-1999
DEFINITION  V.alginolyticus bolA, nqrA, nqrB, nqrC, nqrD and nqrE (partial)
            genes.
ACCESSION   Z37111
VERSION     Z37111.1  GI:663268
KEYWORDS    bolA gene; BolA protein; NADH ubiquinone oxidoreductase;
            NADH:ubiquinone oxidoreductase subunit A; NADH:ubiquinone
            oxidoreductase subunit B; nqrA gene; nqrB gene; nqrC gene; nqrD
            gene; nqrE gene.
SOURCE      Vibrio alginolyticus.
  ORGANISM  Vibrio alginolyticus
```

```
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 2384)
  AUTHORS   Beattie,P., Tan,K., Bourne,R.M., Leach,D., Rich,P.R. and Ward,F.B.
  TITLE     Cloning and sequencing of four structural genes for the
            Na(+)-translocating NADH-ubiquinone oxidoreductase of Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 356 (2-3), 333-338 (1994)
  MEDLINE   95104445
REFERENCE   2  (bases 1 to 5078)
  AUTHORS   Ward,B.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1994) Ward B. F., Edinburgh University, Institute
            for Cell and Molecular Biology, Mayfield Rd, Edinburgh, Scotland,
            EH9 3JR
FEATURES             Location/Qualifiers
     source          1. .5078
                     /organism="Vibrio alginolyticus"
                     /db_xref="taxon:663"
                     /clone="10-2"
                     /clone_lib="Lambda library"
     gene            26. .334
                     /gene="bolA"
     CDS             26. .334
                     /gene="bolA"
                     /function="Cell division"
                     /codon_start=1
                     /transl_table=11
                     /product="BolA"
                     /protein_id="CAA85475.1"
                     /db_xref="GI:663269"
                     /db_xref="SWISS-PROT:Q56585"
                     /translation="MIQEIIEKKLHSELQPSYLKVINESYMHNVPPGSESHFKVIVVS
                     DSFAGQRLIGRHRQVHQILADELDNHIHALAIHTYTDEEWKREQDGAPDSPMCMGGGH
                     "
     gene            808. .2148
                     /gene="nqrA"
     CDS             808. .2148
                     /gene="nqrA"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="NADH:ubiquinone oxidoreductase subunit A"
                     /protein_id="CAA85476.1"
                     /db_xref="GI:663270"
                     /db_xref="SPTREMBL:Q56586"
                     /translation="MITIKKGLDLPIAGTPSQVINDGKTIKKVALLGEEYVGMRPTMH
                     VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEVNRGAKRVLQSVVIEVAGEEQVTF
                     DKFEAAQLSGLDREVIKTQLVDSGLWTALRTRPFSKVPAIESSTKAIFVTAMDTNPLA
                     AKPELIINEQQEAFIAGLDILSALTEGKVYVCKSGTSLPRSSQSNVEEHVFDGPHPAG
                     LAGTHMHFLYPVNAENVAWSINYQDVIAFGKLFLTGELYTDRVVSLAGPVVNNPRLVR
                     TVIGASLDDLTDNELMPGEVRVISGSVLTGTHATGPHAYLGRYHQQVSVLREGREKEL
                     FGWAMPGKNKFSVTRSFLGHVFKGQLFNMTTTTNGSDRSMVPIGNYERVMPLDMEPTL
                     LLRDLCAGDTDSAQALGALELDEEDLALCTFVCPGKYEYGTLLRECLDTIEKEG"
     gene            2118. .3398
                     /gene="nqrB"
     CDS             2118. .3398
                     /gene="nqrB"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase subunit B"
                     /protein_id="CAA85477.1"
                     /db_xref="GI:663271"
                     /db_xref="SPTREMBL:Q56587"
                     /translation="MPRYYREGRVIFMALKKFLEDIEHHFEPGGKHEKWFALYEAVAT
                     VFYTPGIVTNKSSHVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGD
                     QLATVISGNWHYWLTEMLGGTIAADAGVGSKMLLGATYFLPIYATVFLVGGFWEVLFC
                     MVRKHEVNEGFFVTSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPAL
                     AGRAFLFFAYPAQISGDVVWTAVDGFSGATALSQWAQGGNGALVNTVTGSPITWMDAF
                     IGNIPGSIGEVSTLALMIGAAMIVYMRIASGRIIAGVMIGMIAVSTLFNVIGSDTNPM
                     FNMPWHWHLVLGGFAFGMFFMATDPVSASFTNKGKWWYGILIGAMCVMIRVVNPAYPE
                     GMMLAILFANLFAPLFDHVVIEKNIKRRLARYGK"
     gene            3388. .4158
                     /gene="nqrC"
     CDS             3388. .4158
                     /gene="nqrC"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="NADH:uniquinone oxidoreductase"
                     /protein_id="CAA85478.1"
                     /db_xref="GI:663272"
                     /db_xref="SPTREMBL:Q56582"
                     /translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDKQKANAVLD
                     KQSKIVEVAGIDANGKKVPELFAEYIEPRLVDLETGNFTEGNASTYDQREASKDAERS
                     IALTPEEDVADIRRRANTAVVYLVKDQDEVQKVILPMHGKGLWSMMYAFVAVETDGNT
                     VSAITYYEQGETPGLGGEVENPSRRDQFIGKKLYNEDHQPAIKVVKGGAPQGSEHGVD
                     GLSGATLTSNGVQHTFDFWLGDKGFGPFLAKVRDGELN"
     gene            4158. .4790
                     /gene="nqrD"
     CDS             4158. .4790
                     /gene="nqrD"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:uniquinone oxidoreductase"
                     /protein_id="CAA85479.1"
                     /db_xref="GI:663273"
                     /db_xref="SPTREMBL:Q57095"
                     /translation="MSSAQNVKKSILAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                     FGLEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
     gene            4798. .5078
                     /gene="NqrE"
     CDS             4798. .>5078
                     /gene="NqrE"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase"
                     /protein_id="CAA85480.1"
                     /db_xref="GI:663274"
                     /db_xref="SWISS-PROT:Q56589"
                     /translation="MEHYISLLVKSISSKHALSFFLGMCTFLAVSKKVKTSFGLGVAV
                     VVVLTIAVPVNNLVYNLVLRENALVEGVDLSFLNFITFIGVIAALVQIL"
BASE COUNT     1366 a   1068 c   1207 g   1437 t
ORIGIN


  Query Match             35.5%;  Score 477.4;  DB 1;  Length 5078;
  Best Local Similarity   59.7%;  Pred. No. 1.1e-105;
  Matches  803;  Conservative    0;  Mismatches  538;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| ||| || || ||  |  | || || ||||| || |  ||    || || |||
Db     808 ATGATTACAATAAAGAAGGGCTTGGATCTTCCTATCGCAGGAACTCCATCCCAGGTGATT 867

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            |||| ||      ||| |   ||||||| ||||||||||||||| || |  || ||||
Db     868 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 927

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||    ||| | |||   |  || ||||  || |||||||  ||||| || ||||||||
Db     928 CCTACCATGCATGTCCGCGTTGGTGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAT 987

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
           |||||| | || ||||||   ||||| || || |   | || ||| |   ||   | |
Db     988 AAAAAGAACCCTGGCGTGAAATTTACTGCACCAGCAGCCGGTAAAGTGATCGAAGTTAAC 1047

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
           || ||||  || || || ||||| || || |||||||  || | |||  | ||  |  |
Db    1048 CGTGGCGCTAAACGTGTCCTTCAATCTGTAGTGATTGAAGTGGCAGGTGAAGAGCAGGTG 1107

Qy     301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
             || ||    | ||   |||       |  ||    ||    | || |   |
Db    1108 ACATTTGATAAGTTCGAAGCCGCTCAACTTTCAGGTCTAGATCGTGAAGTGATCAAGACT 1167

Qy     361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
```

```
Db   1168 CAGTTGGTTGACTCTGGCCTATGGACCGCTTTACGTACTCGTCCGTTTAGCAAGGTTCCT 1227

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
Db   1228 GCAATCGAGTCTTCTACTAAGGCTATTTTCGTAACTGCAATGGATACTAATCCATTAGCA 1287

Qy    481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
Db   1288 GCTAAGCCTGAGTTGATCATTAATGAGCAACAAGAAGCATTCATTGCTGGTCTAGACATT 1347

Qy    541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
Db   1348 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAA---ATCTGGCACTAGTTTGCCA 1404

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
Db   1405 CGCTCTTCACAGTCTAACGTAGAAGAACACGTCTTCGATGGCCCTCACCCAGCAGGTCTT 1464

Qy    661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
Db   1465 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCATGGAGCATC 1524

Qy    721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacacgcgag 780
Db   1525 AACTACCAAGACGTTATCGCGTTCGGTAAGCTGTTCCTAACTGGTGAGCTATACACAGAT 1584

Qy    781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
Db   1585 CGTGTTGTGTCTCTAGCTGGTCCAGTTGTGAATAACCCTCGTCTAGTTCGTACCGTAATC 1644

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
Db   1645 GGTGCAAGCTTGGATGACCTAACAGATAACGAGTTGATGCCTGGCGAAGTTCGTGTGATT 1704

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
Db   1705 TCTGGTTCTGTACTGACAGGTACACATGCAACTGGTCCACACGCATACCTTGGCCGTTAC 1764

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db   1765 CATCAACAGGTTTCTGTATTACGTGAAGGCCGTGAGAAAGAACTATTTGGCTGGGCAATG 1824

Qy   1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
Db   1825 CCTGGTAAGAATAAGTTCTCGGTAACTCGCTCTTTCCTTGGTCACGTATTCAAAGGTCAG 1884

Qy   1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
Db   1885 TTGTTCAATATGACAACGACGACCAATGGTAGTGATCGTTCAATGGTTCCAATTGGCAAC 1944

Qy   1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
Db   1945 TACGAACGCGTAATGCCTCTAGATATGGAGCCTACTCTGCTGCTTCGTGATCTATGTGCA 2004

Qy   1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
Db   2005 GGTGATACAGACAGTGCTCAAGCACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 2064

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
Db   2065 TTGTGTACCTTTGTATGTCCAGGTAAGTACGAGTACGGTACGTTACTTCGTGAATGCCTA 2124

Qy   1321 gaaaccnttgagaaggaaggctga 1344
Db   2125 GATACTATCGAGAAGGAAGGGTAA 2148


RESULT 14
AB008030
LOCUS       AB008030               6908 bp    DNA     circular BCT 15-JUN-1999
DEFINITION  Vibrio alginolyticus genes for Na-translocating NADH-quinone
            reductase complex, nqr operon, complete genome.
ACCESSION   AB008030 D43958 D49364
VERSION     AB008030.1  GI:2558472
KEYWORDS    Nqr1 subunit of Na-translocating NADH-quinone reductase complex
            alpha-subunit; Nqr2 subunit of Na-translocating NADH-quinone
            reductase complex; Nqr3 subunit of Na-translocating NADH-quinone
            reductase complex gamma-subunit; Nqr4 subunit of Na-translocating
            NADH-quinone reductase complex; Nqr5 subunit of Na-translocating
            NADH-quinone reductase complex; Nqr6 subunit of Na-translocating
            NADH-quinone reductase complex beta-subunit.
SOURCE      Vibrio alginolyticus DNA.
  ORGANISM  Vibrio alginolyticus
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (sites)
  AUTHORS   Hayashi,M., Hirai,K. and Unemoto,T.
  TITLE     Cloning of the Na(+)-translocating NADH-quinone reductase gene from
            the marine bacterium Vibrio alginolyticus and the expression of the
            beta-subunit in Escherichia coli
  JOURNAL   FEBS Lett. 356 (2-3), 330-332 (1994)
  MEDLINE   95104444
REFERENCE   2  (sites)
  AUTHORS   Beattie,P., Tan,K., Bourne,R.M., Leach,D., Rich,P.R. and Ward,F.B.
  TITLE     Cloning and sequencing of four structural genes for the
            Na(+)-translocating NADH-ubiquinone oxidoreductase of Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 356 (2-3), 333-338 (1994)
  MEDLINE   95104445
REFERENCE   3  (sites)
  AUTHORS   Hayashi,M., Hirai,K. and Unemoto,T.
  TITLE     Sequencing and the alignment of structural genes in the nqr operon
            encoding the Na(+)-translocating NADH-quinone reductase from Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 363 (1-2), 75-77 (1995)
  MEDLINE   95246889
REFERENCE   4  (sites)
  AUTHORS   Hayashi,M., Nakayama,Y. and Unemoto,T.
  TITLE     Existence of Na+-translocating NADH-quinone reductase in
            Haemophilus influenzae
  JOURNAL   FEBS Lett. 381 (3), 174-176 (1996)
  MEDLINE   96176316
REFERENCE   5  (sites)
  AUTHORS   Nakayama,Y., Hayashi,M. and Unemoto,T.
  TITLE     Identification of six subunits constituting Na+-translocating
            NADH-quinone reductase from the marine Vibrio alginolyticus
  JOURNAL   FEBS Lett. 422 (2), 240-242 (1998)
  MEDLINE   98149659
REFERENCE   6  (bases 1 to 6908)
  AUTHORS   Hayashi,M., Unemoto,T. and Sugiyama,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-1997) Maki Hayashi, Faculty of Pharmaceutical
            Sciences, Chiba University, Laboratory of Membrane Biochemistry;
            1-33 Yayoi-cho, Inage-ku, Chiba, Chiba 263, Japan
            (E-mail:makiha@p.chiba-u.ac.jp, Tel:043-290-2932, Fax:043-290-3021)
COMMENT     On Oct 22, 1997 this sequence version replaced gi:677955 gi:893412.
            D43958: submitted (17-Dec-1994)
            D49364: submitted (17-Feb-1995).
FEATURES             Location/Qualifiers
     source          1. .6908
                     /organism="Vibrio alginolyticus"
                     /db_xref="taxon:663"
                     /note="There are conflicts between this sequence and
                     GenBank Accession Number:Z37111."
     gene            386. .1726
                     /gene="nqr1"
     CDS             386. .1726
                     /gene="nqr1"
                     /note="contains FMN as a cofactor; nqrA, Nqr1 subunit of
                     Na-NQR complex alpha-subunit"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr1 subunit of Na-translocating NADH-quinone
                     reductase complex alpha-subunit"
                     /protein_id="BAA22910.1"
```

```
                     /db_xref="GI:2558473"
                     /translation="MITIKKGLDLPIAGTPSQVINDGKTIKKVALLGEEYVGMRPTMH
                     VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEVNRGAKRVLQSVVIEVAGEEQVTF
                     DKFEAAQLSGLDREVIKTQLVDSGLWTALRTRPFSKVPAIESSTKAIFVTAMDTNPLA
                     AKPELIINEQQEAFIAGLDILSALTEGKVYVCKSGTSLPRSSQSNVEEHVFDGPHPAG
                     LAGTHMHFLYPVNAENVAWSINYQDVIAFGKLFLTGELYTDRVVSLAGPVVNNPRLVR
                     TVIGASLDDLTDNELMPGEVRVISGSVLTGTHATGPHAYLGRYHQQVSVLREGREKEL
                     FGWAMPGKNKFSVTRSFLGHVFKGQLFNMTTTTNGSDRSMVPIGNYERVMPLDMEPTL
                     LLRDLCAGDTDSAQALGALELDEEDLALCTFVCPGKYEYGTLLRECLDTIEKEG"
conflict             1205
                     /gene="nqr1"
                     /citation=[2]
                     /replace="c"
conflict             1234
                     /gene="nqr1"
                     /citation=[2]
                     /replace="g"
conflict             1256
                     /gene="nqr1"
                     /citation=[2]
                     /replace="t"
gene                 1732. .2976
                     /gene="nqr2"
CDS                  1732. .2976
                     /gene="nqr2"
                     /note="hydrophobic membrane protein with pI 8.14; nqrB,
                     Nqr2 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr2 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22911.1"
                     /db_xref="GI:2558474"
                     /translation="MALKKFLEDIEHHFEPGGKHEKWFALYEAVATVFYTPGIVTNKS
                     SHVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGDQLATVISGNWHY
                     WLTEMLGGTIAADAGVGSKMLLGATYFLPIYATVFLVGGFWEVLFCMVRKHEVNEGFF
                     VTSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPALAGRAFLFFAYPA
                     QISGDVVWTAADGFSGATALSQWAQGGNGALVNTVTGSPITWMDAFIGNIPGSIGEVS
                     TLALMIGAAMIVYMRIASWRIIAGVMIGMIAVSTLFNVIGSDTNPMFNMPWHWHLVLG
                     GFAFGMFFMATDPVSASFTNKGKWWYGILIGAMCVMIRVVNPAYPEGMMLAILFANLF
                     APLFDHVVIEKNIKRRLARYGK"
conflict             1881
                     /gene="nqr2"
                     /citation=[2]
                     /replace="c"
conflict             2400
                     /gene="nqr2"
                     /citation=[2]
                     /replace="t"
conflict             2417. .2418
                     /gene="nqr2"
                     /citation=[2]
                     /replace="tc"
conflict             2614
                     /gene="nqr2"
                     /citation=[2]
                     /replace="g"
gene                 2966. .3736
                     /gene="nqr3"
CDS                  2966. .3736
                     /gene="nqr3"
                     /note="acidic protein with pI 4.83; nqrC, Nqr3 subunit of
                     Na-NQR complex gamma-subunit"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr3 subunit of Na-translocating NADH-quinone
                     reductase complex gamma-subunit"
                     /protein_id="BAA22912.1"
                     /db_xref="GI:2558475"
                     /translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDKQKANAVLD
                     KQSKIVEVAGIDANGKKVPELFAEYIEPRLVDLETGNFTEGNASTYDQREASKDAERS
                     IALTPEEDVADIRRRANTAVVYLVKDQDEVQKVILPMHGKGLWSMMYAFVAVETDGNT
                     VSAITYYEQGETPGLGGEVENPSWRDQFIGKKLYNEDHQPAIKVVKGGAPQGSEHGVD
                     GLSGATLTSNGVQHTFDFWLGDKGFGPFLAKVRDGELN"
conflict             3295
                     /gene="nqr3"
                     /citation=[2]
                     /replace="c"
conflict             3514. .3515
                     /gene="nqr3"
                     /citation=[2]
                     /replace="ta"
gene                 3736. .4368
                     /gene="nqr4"
CDS                  3736. .4368
                     /gene="nqr4"
                     /note="hydrophobic membrane protein with pI 9.17; nqrD,
                     Nqr4 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr4 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22913.1"
                     /db_xref="GI:2558476"
                     /translation="MSSAQNVKKSILAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                     FGLEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
gene                 4376. .4972
                     /gene="nqr5"
CDS                  4376. .4972
                     /gene="nqr5"
                     /note="hydrophobic membrane protein with pI 7.00; nqrE,
                     Nqr5 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr5 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22914.1"
                     /db_xref="GI:2558477"
                     /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGVA
                     VVVVLTIAVPVNNLVYNLVLRENALVEGVDLSFLNFITFIGVIAALVQILEMVLDRFF
                     PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYNFAESIVYGFGSGVGWMLAIVALAGI
                     REKMKYSDVPPGLRGLGITFITVGLMALGFMSFSGVQL"
conflict             4410
                     /gene="nqr5"
                     /citation=[2]
                     /replace=""
conflict             4419
                     /gene="nqr5"
```

```
Query Match             35.4%;  Score 475.8;  DB 1;  Length 6908;
Best Local Similarity   59.7%;  Pred. No. 2.8e-105;
Matches  802;  Conservative    0;  Mismatches  539;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| ||| || || ||  |  | || || ||||| || |  ||    || || |||
Db     386 ATGATTACAATAAAGAAGGGCTTGGATCTTCCTATCGCAGGAACTCCATCCCAGGTGATT 445

Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            |||| ||      ||| |   ||||||| ||||||||||||||| || |  || |||||
Db     446 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 505

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||    ||| | |||   |  || ||||  || |||||||  ||||| || |||||||||
Db     506 CCTACCATGCATGTCCGCGTTGGTGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAT 565

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
           |||||| | || ||||||   ||||| || || |   | || ||| |   ||   |  |
Db     566 AAAAAGAACCCTGGCGTGAAATTTACTGCACCAGCAGCCGGTAAAGTGATCGAAGTTAAC 625

Qy     241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
```

```
Db    626 CGTGGCGCTAAACGTGTCCTTCAATCTGTAGTGATTGAAGTGGCAGGTGAAGAGCAGGTG 685

Qy    301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360

Db    686 ACATTTGATAAGTTCGAAGCCGCTCAACTTTCAGGTCTAGATCGTGAAGTGATCAAGACT 745

Qy    361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420

Db    746 CAGTTGGTTGACTCTGGCCTATGGACCGCTTTACGTACTCGTCCGTTTAGCAAGGTTCCT 805

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480

Db    806 GCAATCGAGTCTTCTACTAAGGCTATTTTCGTAACTGCAATGGATACTAATCCATTAGCA 865

Qy    481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540

Db    866 GCTAAGCCTGAGTTGATCATTAATGAGCAACAAGAAGCATTCATTGCTGGTCTAGACATT 925

Qy    541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600

Db    926 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAA---ATCTGGCACTAGTTTGCCA 982

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660

Db    983 CGCTCTTCACAGTCTAACGTAGAAGAACACGTCTTCGATGGCCCTCACCCAGCAGGTCTT 1042

Qy    661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720

Db   1043 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCATGGAGCATC 1102

Qy    721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780

Db   1103 AACTACCAAGACGTTATCGCGTTCGGTAAGCTGTTCCTAACTGGTGAGCTATACACAGAT 1162

Qy    781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840

Db   1163 CGTGTTGTGTCTCTAGCTGGTCCAGTTGTGAATAACCCTCGTTTAGTTCGTACCGTAATC 1222

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900

Db   1223 GGTGCAAGCTTAGATGACCTAACAGATAACGAGCTGATGCCTGGCGAAGTTCGTGTGATT 1282

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960

Db   1283 TCTGGTTCTGTACTGACAGGTACACATGCAACTGGTCCACACGCATACCTTGGCCGTTAC 1342

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020

Db   1343 CATCAACAGGTTTCTGTATTACGTGAAGGCCGTGAGAAAGAACTATTTGGCTGGGCAATG 1402

Qy   1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080

Db   1403 CCTGGTAAGAATAAGTTCTCGGTAACTCGCTCTTTCCTTGGTCACGTATTCAAAGGTCAG 1462

Qy   1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140

Db   1463 TTGTTCAATATGACAACGACGACCAATGGTAGTGATCGTTCAATGGTTCCAATTGGCAAC 1522

Qy   1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200

Db   1523 TACGAACGCGTAATGCCTCTAGATATGGAGCCTACTCTGCTGCTTCGTGATCTATGTGCA 1582

Qy   1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260

Db   1583 GGTGATACAGACAGTGCTCAAGCACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 1642

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320

Db   1643 TTGTGTACCTTTGTATGTCCAGGTAAGTACGAGTACGGTACGTTACTTCGTGAATGCCTA 1702

Qy   1321 gaaaccnttgagaaggaaggctga 1344
```

```
Db   1703 GATACTATCGAGAAGGAAGGGTAA 1726


RESULT 15
AE004724/c
LOCUS       AE004724               16903 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 285 of 529 of the complete
            genome.
ACCESSION   AE004724 AE004091
VERSION     AE004724.1  GI:9949083
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 16903)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 16903)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .16903
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            complement(67. .2292)
                     /gene="PA2984"
     CDS             complement(67. .2292)
                     /gene="PA2984"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06372.1"
                     /db_xref="GI:9949084"
                     /translation="MRLGLPALAAGLLALRFLAELPSVGWLLGMAAAGLLLLFGRLYW
                     LGLFLLGFAWACGSAQVALDGRLAAELDGRTLWLEGRVVDLPASGEGVTRFRLEEVSA
                     ARGVLLPERLRLSWRGAPPLMGGERWRLAVTLKRPRGLVNAAGFDYEAWLLAQRVGAV
                     GTVKAGQRLQPASGLDAWRDAWRQRLLAVDAQGRGAALAALVLGDGSGLSAAEWRMFQ
                     DTGTVHLMVISGSHISLLAGMLYGLVAALHRLGCWPRRLPWLPCACALGLLGAWSYGL
                     MAGFEVPVRRACLMVSLALVWRLRFRHLGLLTPLLATLVLVLLVEPLASLQAGFWLSF
                     LAVALLLWLFAGRLGRWRWWAAWGRAQWGMALGLAMPLLALGLPLSLSGAPANLLAVP
                     WIELLVVPPALAGSLLLWVPGLGEGLLRAAGVSLDLLFRALGLLAGLAEAWQPPAAAA
                     WSVALGMLGALCWLAPAGLPLRALGAALLLPALLPSSPPVEWGRAEVRVLDVGQGLAV
                     LVRTREHVLLYDSGARQGAFDMGERVVVPVLRSLDLRRLDGLLLSHADNDHAGGAPTV
                     ASRFPPVWLVSGEPARLPSPLFADSCDERSWSWDGVVFEQWAWAQAGDSNDRSCVLRV
                     EADGEVLLLTGDISRAAEHAWLARQADPRVDWLLAPHHGSRSSSGVAFVLRTRPRHVL
                     VSRGWRNAFGHPHGEVMERYARVAAQVHDTARDGALTFLLGSRGGARRERDSAHFWRE
                     K"
     gene            2423. .2956
                     /gene="PA2985"
     CDS             2423. .2956
                     /gene="PA2985"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06373.1"
```

```
                /db_xref="GI:9949085"
                /translation="MPRRLFKRYMPDPESIRSQKSLRFLGKLTHDPNLWHLNRHSVSR
                AMAAGLFAAFIPMPMQMLLAAGLAVWIRSNLPISVGLVWLTNPITMPPVFYCTYKLGA
                WLMQVPPRTLPDHLTWEWISQELSSLWQPFLLGSVVCGIIAAILGYALTMLYWRWWVQ
                RSWSRRCRQRRERVAAK"
     gene       complement(3044. .4345)
                /gene="PA2986"
     CDS        complement(3044. .4345)
                /gene="PA2986"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG06374.1"
                /db_xref="GI:9949086"
                /translation="MPAVRSRPANPPSTRCIRDMFRPLPFFIGLRYTRAKRRNHYISF
                ISLTSMIGLALGVLAMIIVLSVMNGFQKEMRTRILGMVPHATISAAQPLDDWQTVANA
                ALRHREVVGAAPFAELQGMLSYKGNMLPVLVNGIDPQEERKVSIVGEHIVQGSLDDLK
                PGEFGIVLGEITARRFHVNVGDKLTLIVPEATSAPGGITPRMQRFTIVALFKVGAELD
                NSLALIDIADAGQLLRLQPGQVPSVRLELKDLYQSPQVAAKVVKELGQGFRSSDWTRT
                QGSLFNAMKMEKTMIGLLLLLLIIAVAAFNIIATLIMVVADKRTDIAILRTLGATPRQI
                MAIFMVQGTVIGVIGTVIGGVLGVFAALNITGMIDRIERLVGHKVFSSDVYFINYLPS
                DLQVLDVVLICSAALLMSFLATLYPSWRAARTQPAESLRYE"
     gene       complement(4358. .5041)
                /gene="PA2987"
     CDS        complement(4358. .5041)
                /gene="PA2987"
                /codon_start=1
                /transl_table=11
                /product="probable ATP-binding component of ABC
                transporter"
                /protein_id="AAG06375.1"
                /db_xref="GI:9949087"
                /translation="MNDKSVLSCRNLSKSYDEGPQSVQVLSGVELNLLPGERVAIVGS
                SGSGKSTLLNMLGGLDTPSAGSVWLAGEELSALNETARGLLRNRALGFVYQFHHLLPE
                FTALENVCMPLLIGRTPIAEARQRAAELLERVGLGHRLSHKPAELSGGERQRVAIARA
                LVNRPQLVLLDEPTGNLDQHTAQGIQELMLELSRSLQTSFLVVTHDLQLAGHMDRILR
                LEEGRLIAA"
     gene       complement(5034. .6284)
                /gene="PA2988"
     CDS        complement(5034. .6284)
                /gene="PA2988"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG06376.1"
                /db_xref="GI:9949088"
                /translation="MFRPLSVFIGTRYTRAKRRSHFVSFISLTSMIGLALGVLVMIVV
                LSVMNGFDREMRTRILGMVPQATVESYQPIDDWRALAEKVKAHEHVTAVAPFTQMQGM
                LSANGAVQTVMVDAVDPLEEAKVSIIPDFFKQGSLAELKAGGFGIVIGQLAAQKLGVG
                IGDKVTLILPEVAVTPAGVFPRMKRFTVVGTFRVGAGELDGGLSLIHLEDAARLQRWK
                TNQVQGLRLKLDDLFQAPRVAWEIARTLTDNDFYARDWTRSHGNLYQAIRMEKTMIGL
                LLLLIVAVAAFNIISTLVMVVTDKKSDIAILRTLGATPGQIMATFMVQGTVIGVIGTL
                VGGVLGVVAALNVSAWISALEKLLGHQFLASDVYFIDYLPSQLMLDDVVLVCGAALVL
                SFFATLYPAWRAARTQPAEALRYE"
     gene       6222. .6986
                /gene="PA2989"
     CDS        6222. .6986
                /gene="PA2989"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG06377.1"
                /db_xref="GI:9949089"
                /translation="MAATFGPRIARADENGQGSEHVALSWQKMDASGWRSGEASPAAL
                HSDHHCRRGFAMSTRDADDRREYYRIEDTLALEFRPAGAVEDGLESEPGATLFSLLSD
                LHLMDYESQHLLRHISERDRTLANYLKVMNKRIDLLGQVMVQSLLKEIGEPRKVSLSE
                GGVSFRHDRALPVGQLLVLRMVLLPQGFGLELRARVIHAQPHDDEFEIGTEFEALSDA
                QRQLLARHILQKQAQQRRLARAGQGPLGEPGQPSST"
     gene       7318. .8040
                /gene="PA2990"
     CDS        7318. .8040
                /gene="PA2990"
                /codon_start=1
                /transl_table=11
                /product="probable phosphodiesterase"
                /protein_id="AAG06378.1"
                /db_xref="GI:9949090"
                /translation="MTLIYGHRGAKGEAPENTLASFQRCLEHGVNRCELDLHLSRDGE
                LMVIHDPTLKRTTGRRGKVAEHDAATLATYDARKGGPGWVRPTPIPRLEELFEKCPFE
                HWQLEVKSASRERAARTVTRIRELTERFGNRDKVTVTSGSRTVLKALNELTPELSRGL
                VAEYAWLDPLKVAQHYGCALLALNWTLCTPERLLKAQRQGLHVSVWTVNEPALMRRLA
                DFGVDSLITDFPGLASATLGNR"
     gene       complement(8083. .9477)
                /gene="sth"
                /note="PA2991"
     CDS        complement(8083. .9477)
                /gene="sth"
                /codon_start=1
                /transl_table=11
                /product="soluble pyridine nucleotide transhydrogenase"
                /protein_id="AAG06379.1"
                /db_xref="GI:9949091"
                /translation="MAVYNYDVVILGTGPAGEGAAMNASKYGRKLAVVDSRRVVGGNC
                THLGTIPSKALRHSVKQIIEFNTNPMFRQIGEPRWFSFPDVLKSADRVISKQVASRTG
                YYARNRIDMFTGTASFVDERTVEVVTPSGAVERLVADQFVIATGSRPYRPSDINFNHP
                RVYDSDTILSLSHTPRRLIIYGAGVIGCEYASIFSGLGVLVDLIDTRDQLLSFLDDEI
                SDALSYHLRNNNVLIRHNEEYERVEGLDNGVILHLKSGKKIKADALLWCNGRTGNTDK
                LGLENVGIKVNSRGQIEVDENYRTSVSNIFAAGDVIGWPSLASAAYDQGRSAAGNIVE
                SDSWRFVNDVPTGIYTIPEISSIGKNESELTAAKIPYEVGKAFFKGMARAQISNEPVG
                MLKILFHRETLEILGVHCFGDQASEIVHIGQAIMNQPGELNTLKYFVNTTFNYPTMAE
                AYRVAAFDGLNRLF"
     gene       complement(9656. .9883)
                /gene="PA2992"
     CDS        complement(9656. .9883)
                /gene="PA2992"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG06380.1"
                /db_xref="GI:9949092"
                /translation="MTWLLVFVIMLLVVFGMSIGVIMGRKPIAGSCGGIANLGIEKEC
                SICGGDRQKCEEANRDESVVPGKLAYDATKR"
     gene       complement(9880. .10908)
                /gene="PA2993"
     CDS        complement(9880. .10908)
                /gene="PA2993"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG06381.1"
                /db_xref="GI:9949093"
                /translation="MAERFRPLAMAGLAALALVLAGCGDTLESFGGPTMGSTYSIKYV
                RGASAPDVQTAKAAVEAILAEVDRQMSTYRDDSLVSRFNALPAQSCMELPLMLELLR
                YGGELSEQSQGAFDMTVEPLMNLWGFGPQARVEKVPSAEQIAAVRRDVGHRHLRIDGQ
                RLCKDAAVQLDFDSIAAGYTVDAVGERLKELGVRSYLAEITGELKAEGRKPDGTPWRI
                AIEAPREGQRVAQQVLALDGYGVSTSGDYRNYFEENGQRYSHTFDPRTGAPIDHHLAS
                VTVIDPSTRNADGLSTLLMVLGPEEGYRFAEKHRLAALFVSRQGNGFDSRTTPRYEQL
                FGNQGERP"
     gene       complement(10901. .12124)
                /gene="nqrF"
                /note="PA2994"
     CDS        complement(10901. .12124)
```

```
  Query Match             35.2%;  Score 473;  DB 1;  Length 16903;
  Best Local Similarity   59.6%;  Pred. No. 1.4e-104;
  Matches  801;  Conservative    0;  Mismatches  537;  Indels    6;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||| || || |||   || ||  ||||||||||   |||   || || ||   |||
Db  16744 ATGATCAAGATAAAACGTGGCCTGGACCTGCCCATCAGCGGCGCCCCCGAACAGCGCATC 16685

Qy     61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           | |  | | |    |   |   || |||||| | ||    || ||   ||| |||
Db  16684 GAGGCAGCACGCCCGGTACGCAGCGTGGCGTTGATCGGTTTCGACTACCACGGCATGAAG 16625

Qy    121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
```

```
Db  16624 CCGACCATGGCGGTGCAGGTAGGCGACCGGGTGAAGCTGGGGCAGGTGCTGTTCACCGAC 16565

Qy    181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
Db  16564 AAGAAGAATCCCTCCGTCAGCTATACCGCACCCGGCGCCGGTGTGGTCAGTGCCATCCAT 16505

Qy    241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
Db  16504 CGTGGCGAGAAGCGCGTCCTGCAATCGGTGGTCATCGATCTCGACGGCGACGAGCAACTG 16445

Qy    301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
Db  16444 GAGTTCGCCCGTTATCCCGCTGACAAACTCGCCACGCTGAGCGCCGAGCAGGTCCGCGAC 16385

Qy    361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
Db  16384 AACCTGTTGCAGTCGGGCCTGTGGACCGCGCTGCGGACCCGGCCGTTCAGCAAGGTGCCC 16325

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcg 480
Db  16324 GATCCCGAGTCCAGCCCCAGTTCGATTTTCGTCACTGCCATCGATACCCAGCCGCTGGCC 16265

Qy    481 gcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggta 540
Db  16264 GCCGATCCGCAGGTGGTTATCGCGGAGCAGGGCGAGGCCTTCCAGGCCGGGCTGACGGTG 16205

Qy    541 ttgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccg 600
Db  16204 CTCGGGCGCCTGGCCCGG------GTGTTCCTGTGCAAGGCGGAAGGCGTTTCGCTACCC 16151

Qy    601 tctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttg 660
Db  16150 GGCGAGGCGCTCTCCGGCGTGACCGCGCAAGCATTCTCCGGTCCGCATCCGGCGGGCCTG 16091

Qy    661 agtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggaccatc 720
Db  16090 CCCGGCACCCATATCCACTTCCTTGACCCGGTGGGTGCCGGCAAGAGCGTGTGGAACCTG 16031

Qy    721 aattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccgag 780
Db  16030 AACTACCAGGACGTGATCGCCATCGGCAAGCTGTTCACGACCGGGCAATTGTGGACCGAG 15971

Qy    781 cgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttttg 840
Db  15970 CGGGTGATCGCCCTGGCCGGTCCGGTGGTGGAAAAGCCGCGTCTGGTGCGCACCCGCTTG 15911

Qy    841 ggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtgatt 900
Db  15910 GGCGCCAACCTCGACGAACTGGCCGCCGGCCAGTTGCAGCCGGGCAACAACCGGCTGATC 15851

Qy    901 tccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgggacgctac 960
Db  15850 TCCGGCTCCGTGCTGGGCGGTCGCACCGCCCATGGCGCCTACGCCTACCTGGGGCGCTAC 15791

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db  15790 CACCTGCAGTTGTCCTGTCTCAAGGAGGGCGACCAGCGCGAGTTCCTCCACTACCTGCGC 15731

Qy   1021 ccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaaacaaa 1080
Db  15730 GCCGGGGTGGAAAAGCATTCGCTGCTCAACGTGTTCGTCTCCAGGCTGCTCGGCGGCAAG 15671

Qy   1081 ctcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattggtact 1140
Db  15670 CGCTTCGCCTTCACCACCAGCACCAACGGCAGCCCACGGGCGATGGTGCCGGTAGGTAAC 15611

Qy   1141 tacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgatttaatcgtc 1200
Db  15610 TACGAGGCGGTGATGCCGCTGGATATCCTGCCGACTCAATTGCTGCGCTACCTGATCGTC 15551

Qy   1201 ggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaagacctcgct 1260
Db  15550 GGCGACACGGAAATGGCCCAGAAGCTCGGCGCCCTGGAACTCGACGAAGAGGACCTGGCG 15491

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgtaaggtgctg 1320
Db  15490 CTGTGCAGCTACGTCTGCGCCGGCAAGTACGAATACGGCCCGATTCTGCGGGACAATCTG 15431

Qy   1321 gaaaccnttgagaaggaaggctga 1344
Db  15430 GCACGCATCGAGCAGGAGGGCTGA 15407
```

RESULT 16
A96090
LOCUS A96090 474 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 123 from Patent WO9924578.
ACCESSION A96090
VERSION A96090.1 GI:6779937
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 474)
AUTHORS Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
TITLE Neisserial antigens
JOURNAL Patent: WO 9924578-A 123 20-MAY-1999;
PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES Location/Qualifiers
source 1. .474
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 122 a 130 c 130 g 91 t 1 others
ORIGIN

Query Match 32.1%; Score 431.6; DB 6; Length 474;
Best Local Similarity 93.5%; Pred. No. 1.6e-94;
Matches 443; Conservative 1; Mismatches 30; Indels 0; Gaps 0;

```
Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
Db      1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 60

Qy     61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
Db     61 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy    121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
Db    121 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy    181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
Db    181 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 240

Qy    241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatc 300
Db    241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAARGCAACGACGAAATC 300

Qy    301 gagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngc 360
Db    301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 360

Qy    361 aatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccct 420
Db    361 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 420

Qy    421 gccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccg 474
Db    421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG 474
```

RESULT 17

```
AX067460/c
LOCUS       AX067460               96109 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 35 from Patent WO0078968.
ACCESSION   AX067460
VERSION     AX067460.1  GI:12545080
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 96109)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 35 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .96109
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
BASE COUNT    28783 a  18910 c  20341 g  28075 t
ORIGIN
```

Query Match 24.8%; Score 333.6; DB 6; Length 96109;
Best Local Similarity 54.6%; Pred. No. 1.3e-70;
Matches 742; Conservative 0; Mismatches 596; Indels 21; Gaps 4;

```
Qy      1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
          |||||||  ||||||||||  |  | ||||| ||| | |||   ||     |   ||
Db  85201 ATGATTACCATCAAAAAAGGCTTGGATCTGCCAATCACTGGCGAGCCCTCCAAGCAGAT- 85141

Qy     61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
            |  || || |      ||  | || |||||  | ||| | || |||| ||| |||||
Db  85142 --TAGCGAGCACACGCCCACTAAGGTGGCGTTAGTCGGCTACGATTATGTCGGCATGCGT 85085

Qy    121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||    ||| | ||||| ||||| |||   ||||  ||||| || |||||  |  |||||||
Db  85084 CCGACCATGCATGTCAAAGAAGGTGATAAGGTCACCAAAGGACAGGTGCTTTACGAAGAC 85025

Qy    181 aaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccat 240
          |||||     | |||||  | | ||| || || |||   ||||  |||  ||  || ||
Db  85024 AAAAAACGAGCAGGCGTCCTTTATACAGCACCAGTTGGCGGCACTGTCGTTGCTGTCAAT 84965

Qy    241 cgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttg------aaggcaacgac 294
          || ||||||   || ||| || |     | || || ||||| |      | |  |  ||
Db  84964 CGTGGCGAACGACGAGTATTTGAAAGCTTGGTCATCGCCGTCGATCCAAATGCAAGTGAA 84905

Qy    295 gaaatcgagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantn 354
          ||||||   ||  |    ||     | ||||  ||| |||   |     || |  | |
Db  84904 GAAATCACTTTTAATGCTTATCGCTCTGAAGATTTGTCAATGCTGGATCGCAAGACAGTT 84845

Qy    355 ngnngcaatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaa 414
              |  || ||   || |||  ||||| || || || |  |||||| || |||
Db  84844 GTTGAGCAGTTGGTCGCTTCAGGTGAATGGACCGCCCTACGCACTCGTCCTTTTAGCCGT 84785

Qy    415 atccctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccg 474
          |  || |   |  ||  |   || |  ||||| || || |  || |  || |||||||||
Db  84784 ACACCAGAAATTAATAGCATACCTTCTGCCATTTTTGTGACGGCAACAGATACCAATCCT 84725

Qy    475 ctngcggcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntg 534
           | || |  |||||||     ||  |    ||||      | |  || |       |  ||
Db  84724 TTGGCTGTTGACCCTGCCACCGTATTATTTGAAGAAATACAAGCATTTAATGATGGCTTG 84665

Qy    535 ctggtattgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagac 594
            | ||        |  ||   |  |||| | |||| ||| | |  | |   | |
Db  84664 GCAATTTTATCAACGCTATCC---CCAAAAACCTATGTCTGTCATGGTGATATTGTACCT 84608

Qy    595 gtgccgtctgaaaatgctgccaac---------atcgaaacacatgaattcggcggcccg 645
          | | | |     ||  | ||  ||         | |       | |  ||    ||
Db  84607 GAGACCTTAAGTAAAACCGCTGACGGTAACGATACCATCTATGAAGGGTTTACTGGTGTA 84548

Qy    646 catccggccggtttgagtggcacgcacattcatttcattgagccggtcggtgcaaacaaa 705
          |||||  || ||       || ||||||||||| ||  |  | ||  | |    |  ||
```

```
Db  84547 CATCCAGCAGGCAACGCAGGTACGCACATTCACTTTTTACACCCTTTGGCACGAGGCAGC 84488

Qy    706 accgtttggaccatcaattatcaagatgtaattgccatcggacgtttgtttgcaacaggc 765
            |||||||| || |||||||||||||| || ||||| ||     | ||| | |||||
Db  84487 TTAGTTTGGACGATTAATTATCAAGATGTCATCGCCATTGGTAAGCTATTTACCACAGGT 84428

Qy    766 cgtctgaacaccgagcgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctc 825
          ||  |  |||| |||||  ||||   |||||  ||| |   |||   ||| || ||| |
Db  84427 CGCATTTACACAGAGCGTCTGATCAGTTTGGCAGGTCCAGCAGTGGCCAATCCTCGCTTG 84368

Qy    826 ttgcgtaccgttttgggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgca 885
            ||||||||  |  |||||  |  |  |  || | ||    || ||| |    | ||
Db  84367 GTGCGTACCACTCGAGGTGCTGATTTGACCGAATTGACCCAAGGTGAACTCAACGGCGAC 84308

Qy    886 gacaaccgcgtgatttccggttcggtattgaacggcgcgattacacaaggcgcgcacgat 945
          || ||||| || ||||| || || ||  |    ||    |   |    || ||    |
Db  84307 GATAACCGTGTCATTTCAGGCTCTGTTCTATCAGGTCGCACAGCTGTTGGTGCTACGGCA 84248

Qy    946 tatttgggacgctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctg 1005
          | |||||| ||||  || |||||  | || ||  |    |||| ||
Db  84247 TTTTTGGGCCGCTTTCATAATCAAGTCTCTGTACTTACCGAAGGTCGTGAACGCCCTGCA 84188

Qy   1006 ttcggctgggttgcgccgcagccggacaaatactccatcacgcgtacgaccctcggccat 1065
            |  ||   || | ||   | |  |    |  || |    ||  |      |   ||
Db  84187 CTTCACTTTTTTACACCAGGGGCAAATCGTTTTTCAAAATTGCCAATTTATATTTCACAG 84128

Qy   1066 ttcctgaaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatg 1125
          ||  | |||  ||||    | | | || |||||      ||||||       || ||||||
Db  84127 TTTTTTAAAGGCAAAAAATACGATTTTACGACCAGTACCAACGGCTCGCCTCGTGCCATG 84068

Qy   1126 gtgccgattggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttg 1185
          || || || ||    |  ||    || ||||| | ||||   || |||||| ||| ||
Db  84067 GTACCAATCGGCTTATTTGAAACTGTCATGCCTCAAGACTATCTACCTACCCAGCTACTG 84008

Qy   1186 cgcgatttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggac 1245
          || | ||| || ||||  |||| |  ||| |||   | |||||||    | ||  | ||
Db  84007 CGTGCTTTGATTGTCGAAGATATCATCAGTGCGGTTGAATTGGGTGCACTTGAGCTTGAT 83948

Qy   1246 gaagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctg 1305
          ||||||||  | ||  |||| |  ||||| |  || |||||||| || |  ||     |
Db  83947 GAAGAAGATTTGGCACTGTGTACTTTCGTATCGCCAGGCAAATATGAGTTCGGTGATATT 83888

Qy   1306 ttgcgtaaggtgctggaaaccnttgagaaggaaggctga 1344
          ||||||  |      ||       ||||| |||||||| |
Db  83887 TTGCGTGATAACTTGACCCGTATTGAGCAGGAAGGCTAA 83849
```

```
RESULT  18
AF144076
LOCUS       AF144076               1356 bp    DNA     linear   BCT 01-JUN-1999
DEFINITION  Porphyromonas gingivalis strain W50 50 kD antigen PG1 gene,
            complete cds.
ACCESSION   AF144076
VERSION     AF144076.1  GI:4929279
KEYWORDS    .
SOURCE      Porphyromonas gingivalis.
  ORGANISM  Porphyromonas gingivalis
            Bacteria; CFB group; Bacteroidetes; Bacteroidales;
            Porphyromonadaceae; Porphyromonas.
REFERENCE   1  (bases 1 to 1356)
  AUTHORS   Ross,B.C., Barr,I., Patterson,M., Agius,C., Rothel,L., Margetts,M.,
            Hocking,D. and Webb,E.
  TITLE     Porphyromonas gingivalis polypeptides and nucleic acids
  JOURNAL   Patent: Australia (AU 98/01023)-PCT 10-DEC-1998;
            CSL Ltd;
            45 Poplar Road;
            R&D;
            Parkville, Victoria;
            Australia;
REFERENCE   2  (bases 1 to 1356)
  AUTHORS   Ross,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1999) R&D, CSL Ltd, 45 Poplar Road, Parkville,
```

```
           Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .1356
                     /organism="Porphyromonas gingivalis"
                     /strain="W50"
                     /db_xref="taxon:837"
     CDS             1. .1356
                     /codon_start=1
                     /transl_table=11
                     /product="50 kD antigen PG1"
                     /protein_id="AAD33930.1"
                     /db_xref="GI:4929280"
                     /translation="MAKVIKTKKGLALNLKGKPLPEMLAEPAQSPTYAVVPDDFEGVI
                     PKVTARPGDKVRAGSALMHHKAYPEMKFTSPVSGEVIAVNRGAKRKVLSIEVKPDGLN
                     EYESFPVGDPSALSAEQIKELLLSSGMWGFIKQRPYDIVATPDIAPRDIYITANFTAP
                     LAPDFDFIVRGEERALQTAIDALAKLTTGKVYVGLKPGSSLGLHNAEIVEVHGPHPAG
                     NVGVLINHTKPINRGETVWTLKATDLIVIGRFLLTGKADFTRMIAMTGSDAAAHGYVR
                     IMPGCNVFASFPGRLTIKESHERVIDGNVLTGKKLCEKEPFLSARCDQITVIPEGDDV
                     DELFGWAAPRLDQYSMSRAYFSWLQGKNKEYVLDARIKGGERAMIMSNEYDRVFPMDI
                     YPEYLLKAIIAFDIDKMEDLGIYEVAPEDFATCEFVDTSKIELQRIVREGLDMLYKEM
                     N"
BASE COUNT      338 a    333 c    380 g    305 t
ORIGIN
```

Query Match 9.1%; Score 122.2; DB 1; Length 1356;
Best Local Similarity 46.0%; Pred. No. 4.2e-19;
Matches 622; Conservative 0; Mismatches 700; Indels 30; Gaps 6;

```
Qy       2 tgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcattt 61
Db      11 TTATAAAAACAAAAAAAGGCCTTGCACTTAATCTGAAAGGAAAACCGCTGCCCGAGATGC 70

Qy      62 atgacgggcccgtc---attaccgaagtcgcgttgcttggcgaagaatatgccggtatgc 118
Db      71 TGGCCGAACCGGCCCAAAGTCCTACTTACGCGGTCGTGCCCGACGATTTTGAAGGTGTTA 130

Qy     119 gccccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaag 178
Db     131 TCCCCAAGGTGACGGCTCGTCCGGGGGATAAGGTGCGTGCCGGCTCAGCACTGATGCACC 190

Qy     179 acaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatcc 238
Db     191 ACAAGGCATATCCGGAGATGAAGTTTACAAGTCCGGTTAGCGGCGAAGTGATCGCGGTGA 250

Qy     239 atcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaa 298
Db     251 ATCGCGGTGCCAAGCGC------AAGGTGTTGAGCATCGAGGTGAAACCGGACGGACTGA 304

Qy     299 tcgagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnn 358
Db     305 ACGAATACGAGTCATTCCCTGTCGGGGATCCGTCTGCCCTCTCTGCCGAACAGATCAAGG 364

Qy     359 gcaatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatcc 418
Db     365 AGCTTTTACTGTCGAGCGGTATGTGGGGTTTTATTAAGCAACGTCCTTACGACATAGTGG 424

Qy     419 ctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctng 478
Db     425 CTACACCGGATATAGCTCCACGCGACATTTATATTACTGCCAACTTTACTGCACCATTGG 484

Qy     479 cggcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctgg 538
Db     485 CTCCGGACTTCGATTTCATCGTTCGAGGAGAAGAACGCGCCCTGCAGACTGCCATCGATG 544

Qy     539 tattgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgc 598
Db     545 CCTTGGCCAAACTCACGACAGGAAAGGTGTATGTGGGCCTGAAGCCGGGTTCATCTCTGG 604

Qy     599 cgtctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtt 658
Db     605 GCTTGCACAATGCAGAAATCGTAGAAGTACA---------CGGACCTCATCCGGCAGGTA 655

Qy     659 tgagtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggacca 718
Db     656 ACGTGGGCGTGCTGATCAATCATACGAAGCCAATCAATCGGGGCGAAACGGTGTGGACGC 715

Qy     719 tcaattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccg 778
Db     716 TCAAGGCTACCGACCTGATCGTGATCGGACGTTTCCTGCTTACGGGCAAAGCCGATTTTA 775

Qy     779 agcgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttt 838
Db     776 CCAGAATGATTGCCATGACCGGCTCAGACGCTGCAGCTCACGGATACGTCCGTATTATGC 835

Qy     839 tgggtgcgaaagtatcgcaaattactgcgggcgaattg------gttgacgcagacaacc 892
Db     836 CGGGTTGCAATGTCTTTGCTTCCTTCCCCGGCCGACTGACAATAAAGGAATCTCACGAGC 895

Qy     893 gcgtgatttccggttcggtattgaacggcgcgattacacaaggcgcgcacgattatttgg 952
Db     896 GTGTGATCGATGGCAATGTGCTGACCGGTAAGAAGCTCTGCGAGAAGGAGCCTTTCCTGT 955

Qy     953 gacgctaccacaatcagatttccgttatcgaagaaggccgc---agcaaagagctgttcg 1009
Db     956 CAGCCCGGTGTGACCAGATCACGGTGATCCCCGAAGGCGACGATGTGGACGAACTCTTCG 1015

Qy    1010 gctgggttgcgccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcc 1069
Db    1016 GGTGGGCTGCACCCCGTCTCGATCAGTACAGCATGAGCAGAGCTTATTTCTCTTGGTTGC 1075

Qy    1070 ---tgaaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatgg 1126
Db    1076 AGGGGAAAAACAAAGAGTACGTACTCGATGCCCGGATCAAGGGTGGCGAACGTGCTATGA 1135

Qy    1127 tgccgattggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgc 1186
Db    1136 TCATGAGCAACGAGTATGACCGCGTTTTCCCGATGGACATCTATCCGGAGTATTTGCTCA 1195

Qy    1187 gcgatttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacg 1246
Db    1196 AGGCTATTATAGCATTCGACATCGACAAGATGGAGGACTTAGGCATATATGAAGTGGCTC 1255

Qy    1247 aagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgt 1306
Db    1256 CGGAGGACTTTGCCACTTGCGAATTTGTGGATACATCCAAGATCGAGCTGCAGCGTATCG 1315

Qy    1307 tgcgtaaggtgctggaaaccnttgagaaggaa 1338
Db    1316 TTCGCGAGGGCTTGGATATGCTCTATAAGGAA 1347
```

```
RESULT 19
HIU20229/c
LOCUS       HIU20229                4926 bp    DNA     linear   BCT 09-FEB-1995
DEFINITION  Haemophilus influenzae BOLA (bolA), glutathione reductase (gor),
            phosphatidylserine decarboxylase (psd), 30K protein (rpmF), genes,
            complete cds.
ACCESSION   U20229
VERSION     U20229.1  GI:644850
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 4926)
  AUTHORS   Barcak,G.J. and Heimer,S.R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4926)
  AUTHORS   Barcak,G.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-1995) Gerard J. Barcak, Biochemistry, University
            of Maryland at Baltimore, 108 North Greene Street, Baltimore, MD
            21201, USA
FEATURES             Location/Qualifiers
     source          1. .4926
```

```
                     /organism="Haemophilus influenzae"
                     /strain="Rd"
                     /isolate="KW20"
                     /db_xref="taxon:727"
     CDS             complement(27. .224)
                     /note="orf65"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62134.1"
                     /db_xref="GI:644851"
                     /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                     VREGDVVKKGQVLFEDKKILV"
     gene            complement(519. .842)
                     /gene="bolA"
     CDS             complement(519. .830)
                     /gene="bolA"
                     /function="stationary phase regulatory protein"
                     /codon_start=1
                     /transl_table=11
                     /product="BOLA"
                     /protein_id="AAA62135.1"
                     /db_xref="GI:644852"
                     /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                     SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                     R"
     misc_feature    775. .785
                     /note="encodes DNA uptake recognition sequence"
     RBS             complement(838. .842)
                     /gene="bolA"
     protein_bind    877. .913
                     /note="OxyR recongition sequence (Tartaglia, L.A., et al.,
                     J. Mol Biol. 210: 709-710, 1989)"
                     /bound_moiety="OxyR"
     RBS             915. .919
                     /note="orf199"
     CDS             924. .1523
                     /note="orf199"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62136.1"
                     /db_xref="GI:644853"
                     /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                     VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                     QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                     VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
     gene            1343. .3109
                     /gene="gor"
     -35_signal      1343. .1348
                     /gene="gor"
                     /note="putative"
     -10_signal      1365. .1370
                     /gene="gor"
                     /note="putative"
     -35_signal      1585. .1590
                     /gene="gor"
                     /note="putative"
     -10_signal      1605. .1610
                     /gene="gor"
                     /note="putative"
     RBS             1728. .1731
                     /gene="gor"
     CDS             1739. .3109
                     /gene="gor"
                     /EC_number="1.6.4.2"
                     /function="reduces oxidized glutathione using NADPH"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="glutathione reductase"
                     /protein_id="AAA62137.1"
                     /db_xref="GI:644854"
                     /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                     VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                     LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                     DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                     LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                     LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                     KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                     QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                     EFVTMR"
     gene            complement(3218. .4090)
                     /gene="psd"
     CDS             complement(3218. .4090)
                     /gene="psd"
                     /function="lipid metabolism"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphatidylserine decarboxylase"
                     /protein_id="AAA62138.1"
                     /db_xref="GI:644855"
                     /translation="MMTSNSYWQRLKVAFQYVMPQIYLTQIAGWFAKQKWGKITHFVI
                     KAFAKKYNIDMSIAQKEQFSDYASFNEFFIRPLKENARPINQNPTALCLPADGRISEC
                     GHIDDNLLLQAKGHFFSLEDLLAEDKELVETFKNGEFVTTYLSPRDYHRVHMPCDATL
                     RKMIYVPGDLFSVNPFLAQHVPNLFARNERVICVFDTEFGTMVQILVGATITASIGTT
                     WAGVINPPRHNEVKTWTYEGESAVKLLKGQEMGWFQLGSTVINLFQANQVRLADHLSV
                     NEPVRMGEILAYKK"
     misc_feature    complement(3364. .3374)
                     /gene="psd"
                     /note="encodes DNA uptake recognition sequence"
     misc_feature    complement(3780. .3790)
                     /gene="psd"
                     /note="encodes DNA uptake recognition sequence"
     misc_feature    3818. .3825
                     /note="encodes DNA uptake recognition sequence"
     RBS             complement(4097. .4100)
                     /note="psd"
     protein_bind    complement(4126. .4152)
                     /note="OxyR recognition sequence (Toledano, M.B. et al.,
                     Cell 78:897-909.)"
                     /bound_moiety="OxyR"
     gene            4247. .4771
                     /gene="rpmF"
     CDS             4247. .4771
                     /gene="rpmF"
                     /codon_start=1
                     /transl_table=11
                     /product="30K protein"
                     /protein_id="AAA62139.1"
                     /db_xref="GI:644856"
                     /translation="MQKVKLPLTVDPIKDAQRRLDYVGYYAANQLERLAESVVNVLSD
                     AQVTLSFYVDPQKLVVMKGKVQIDVDLECQRCNEPYKQTLECEFTYSPVANWDQADDL
                     PEIYEPIEFNEFGEIDLIGTVEDELILALPLVPMHSSEHCEVSAQEQVFGVLPEELAK
                     KPNPFAVLANLKQK"
     CDS             complement(4561. .>4926)
                     /note="orf121"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62140.1"
                     /db_xref="GI:644857"
                     /translation="LQLRGGVSEFHRLLYLLIQQQLLKHRVNDAYHDVNEIFCFVEQP
                     LAILLNLIYFCFKLANTANGFGFFANSSGNTPKTCSCADTSQCSDECIGTSGKAKISS
                     SSTVPIKSISPNSLNSIGS"
BASE COUNT     1598 a    862 c    938 g   1528 t
ORIGIN


  Query Match             7.5%;  Score 101.4;  DB 1;  Length 4926;
  Best Local Similarity   69.6%;  Pred. No. 5e-14;
  Matches  149;  Conservative    0;  Mismatches   64;  Indels    1;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaacctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| ||| || |||||| |  | || || || ||||| | ||| |  ||||| ||
Db     224 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 165
```

```
Qy      61 tatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgc 120
           ||  |||   ||   | |   |||| ||| | || || || || ||||  || ||||||
Db     164 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 105

Qy     121 ccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaagac 180
           || |  ||||| ||    ||||||||||  || ||||||||| ||||| || ||||||||||
Db     104 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTGAAGAC 45

Qy     181 aaaaagnatccgggcgtggtgtttaccgcgccng 214
           |||||  |||| || ||  | ||||| || || |
Db      44 AAAAA-AATCCTGGTGTAATTTTTACAGCCCCTG 12


RESULT  20
AF429315/c
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN


  Query Match             3.4%;  Score 45.6;  DB 9;  Length 125020;
  Best Local Similarity  10.7%;  Pred. No. 2.1;

  Matches   97;  Conservative  374;  Mismatches  431;  Indels    7;  Gaps    2;

Qy     339 aagcggcgangaantnngnngcaatctgatccaatccggtttgtggactgcgctgcgtan 398
           | : :: |  |   :  |  :  : ::::: :     :|    :  : ::|:::  :
Db   17784 AGSKSSRGGTGYYMKKKGGGKGSMSSKKWKGSSTSRRRGSSAKSCSYMWGMSMCMSCMSM 17725

Qy     399 ccgtccgttcagcaaaatccctgccgtcgatgccgagccgttcgccatcttcgtcaatgc 458
             :::: : ::  :    :::::|:: :|: : :   :| :::::   |::| ::    :
Db   17724 AKSYMMCYCYMYRMSSYMSYYYKCYSCMGMSSTSYSYSCCWKMSSWGSCYKCMKYYSGWS 17665

Qy     459 gatggacaccaatccgctngcggcagaccctgtggttgtgatcaaagaagccgncganga 518
           :   ::  ::   | ::: : ::: :   ::::::: ::  :    :| : :   |: ::
Db   17664 SYSTSMGYYSSTSCKKYKCSWSSMYKCKTSKSYRRKRSYYYWGGGKRAKKKYYCAGRRRR 17605

Qy     519 tttcagacgangtntgctggtattgagccgtttgaccgagcgtaaaatccatgtgtgtaa 578
             ::: |      : |::  :|:  : : |:::  |  : |: :  : |   |::  :
Db   17604 -MSYWKCCAKWWMSYCCWSYCMTYYYSKSCTYKSSTCYKRGGYYWGSKTCYSAGGKSRSM 17546

Qy     579 ggcagctggcgcagacgtgccgtctgaaaatgctgccaacatcgaaacacatgaattcgg 638
             |:   :::::::::  :::  :   |  :::::  : |  :       ||:   |::
Db   17545 YYCMMRSSKSSSWSMSMARSSWCMGWGAGYRRSKRSAGWGAGWRSSKGKRSTGMKRACSK 17486

Qy     639 cggcccgcatccggccggtttgagtggcacgcacattcatttcattgagccggtcggtgc 698
             |:: :   ::::  |  ::::| :|    | |:::: ::       : :: :  | :::
Db   17485 KTGSYGSTGRSMKKKKGYSKYSRGMKGKKKTCYCMWKYYKYRKTSMCWWYYMKSWGYKRY 17426

Qy     699 aaacaaaaccgtttggaccatcaattatcaagatgtaattgccatcggacgtttgtttgc 758
            : |:    |  : |  :        : :: : ::|  :::|  ::   | :: :::::
Db   17425 KRCCMKKKGCTGYRGMSSKKSGYKSMRGMRGSSYSTSCWKSCWGYSWMKCMKYMSYKKKR 17366

Qy     759 aacaggccgtctgaacaccgagcgcgtgattgctttgggtggttctcaagtcaacaaacc 818
           : ::|::::: ::         ::::   :: ::: :: :  ::    : |:  : :   :
Db   17365 RSMRGSMSSKGWRGYAGRGCYSSSMWSTRKRRSKCYSYKSYKKGRGKMKGWGGMKRGSKY 17306

Qy     819 acgcctcttgcgtaccgttttgggtgcgaaagtatcgcaaattactgcgggcgaattggt 878
           :::: ::: ::   |::|: :::| : :  :|  | : |    :: |::  :|  :::
Db   17305 WSSMKKMKRSSWSKCYSTKSYKSGRRKSKGWGRSTKSKAKSSMRMAGSKCTYGSSYWSNN 17246

Qy     879 tgacgcagacaaccgcgtgatttccggttcggtattgaacggcgcgatta------caca 932
           :  :: :   ::  : :|           |    |   | | ||  |      :|::
Db   17245 RRNNRMGKTGCNYMYRRSRAMMNGNAAAAGCTTCCCCANTNGGGGGAAAAAGGGCGSASR 17186

Qy     933 aggcgcgcacgattatttgggacgctaccacaatcagatttccgttatcgaagaaggccg 992
           |: :: :: ::  ::  |::: ::::    :: :::|:|   :| :  ::| | :||:
Db   17185 ASCYKGRMSSKSCYRSGTRRRCMKSKCRRGSRGKSMGMTRGRSGGKTSYSAKGSGRGCYY 17126

Qy     993 cagcaaagagctgttcggctgggttgcgccgcagccggacaaatactccatcacgcgtac 1052
           |:|  : |::| :    ::::: : : : |:::| ::::|::   :|  |:|::   |
Db   17125 CWGWGRKGRKCMSSRKMKMYKSYYRRKRWMTCMKMCYSMYMAMYCRSMCCMCMCKSCCGC 17066

Qy    1053 gaccctcggccatttcctgaaaaacaaactcttcaagttcacgacagccgtcaacggtgg 1112
           : : : ::: ::    :  ::::     :|  |::   :     | ::   :: |:||
Db   17065 YSMGMSSYSYSGKYSWGMKSYMWRSYYSKRSTSKAWRSSKRGMGTGGRYKGGGRSYGKGG 17006

Qy    1113 cgaccgcgccatggtgccgattggtacttacgagcgcgtaatgccgctagacatcctgcc 1172
             :   | :   :||::  |   |:: ::: ::   ::: : :: | : : :|: :::
Db   17005 GSYKGGGSWGGKGGKSTRGSSAGKKSSYKKMSCCARYKSMSYSKCYSRWMWCMSYYYCWG 16946

Qy    1173 taccctgcttttgcgcgatttaatcgtcggcgataccgacagcgcgcaagcattgggttg 1232
             | :   :  : :::   :|  ::: :|:: |: :: : : ::: : : :  ::: :
Db   16945 GKCWSSCWWSSMRKKSKGGSWAKGMVWDKGSVSTDKSDKMBSRSBSKVKSKWSMSRRGKY 16886

Qy    1233 cttggaatt 1241
           ::  : :::
Db   16885 YYCTDSMKK 16877


RESULT  21
BACGNTZA/c
LOCUS       BACGNTZA               36200 bp    DNA     linear   BCT 05-FEB-1999
DEFINITION  Bacillus subtilis 36kb sequence between gntZ and trnY genes
            encoding 34 ORFs.
ACCESSION   D78193
```

```
VERSION     D78193.1  GI:1064780
KEYWORDS    phospho-bete-glucosidase; 6-phosphogluconate dehydrogenase; alkyl
            hydroperoxide reductase; amino acid permease; arginase; orthinine
            aminotransferase; positive regulatory protein; gntZ; ahp; bglA;
            yyd; yyc; formaldehyde dehydrogenase.
SOURCE      Bacillus subtilis (strain:168) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 36200)
  AUTHORS   Ogasawara,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1995) Naotake Ogasawara, Nara Institute of
            Science and Technology, Graduate School of Biological Sciences;
            8916-5 Takayama-cho, Ikoma, Nara 630-01, Japan
            (E-mail:nogasawa@bs.aist-nara.ac.jp, Tel:07437-2-5430,
            Fax:07437-2-5439)
REFERENCE   2  (sites)
  AUTHORS   Zhang,J. and Aronson,A.
  TITLE     A Bacillus subtilis bglA gene encoding phospho-beta-glucosidase is
            inducible and closely linked to a NADH dehydrogenase-encoding gene
  JOURNAL   Gene 140 (1), 85-90 (1994)
  MEDLINE   94171085
REFERENCE   3  (sites)
  AUTHORS   Hartford,O.M. and Dowds,B.C.
  TITLE     Isolation and characterization of a hydrogen peroxide resistant
            mutant of Bacillus subtilis
  JOURNAL   Microbiology 140 (Pt 2), 297-304 (1994)
  MEDLINE   94236234
REFERENCE   4  (sites)
  AUTHORS   Ogasawara,N., Nakai,S. and Yoshikawa,H.
  TITLE     Systematic sequencing of the 180 kilobase region of the Bacillus
            subtilis chromosome containing the replication origin
  JOURNAL   DNA Res. 1 (1), 1-14 (1994)
  MEDLINE   96051385
REFERENCE   5  (sites)
  AUTHORS   Calogero,S., Gardan,R., Glaser,P., Schweizer,J., Rapoport,G. and
            Debarbouille,M.
  TITLE     RocR, a novel regulatory protein controlling arginine utilization
            in Bacillus subtilis, belongs to the NtrC/NifA family of
            transcriptional activators
  JOURNAL   J. Bacteriol. 176 (5), 1234-1241 (1994)
  MEDLINE   94156824
REFERENCE   6  (sites)
  AUTHORS   Yoshida,K., Seki,S., Fujimura,M., Miwa,Y. and Fujita,Y.
  TITLE     Cloning and sequencing of a 36-kb region of the Bacillus subtilis
            genome between the gnt and iol operons
  JOURNAL   DNA Res. 2 (2), 61-69 (1995)
  MEDLINE   96093926
REFERENCE   7  (sites)
  AUTHORS   Gardan,R., Rapoport,G. and Debarbouille,M.
  TITLE     Expression of the rocDEF operon involved in arginine catabolism in
            Bacillus subtilis
  JOURNAL   J. Mol. Biol. 249 (5), 843-856 (1995)
  MEDLINE   95311309
REFERENCE   8  (bases 1 to 36200)
  AUTHORS   Kasahara,Y., Nakai,S., Yoshikawa,H. and Ogasawara,N.
  TITLE     36kb sequence between gntZ and trnY of B. subtilis genome
  JOURNAL   Unpublished (1997)
FEATURES             Location/Qualifiers
     source          1. .36200
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            1. .32
                     /gene="gntZ"
     CDS             <1. .32
                     /gene="gntZ"
                     /citation=[6]
                     /codon_start=3
                     /transl_table=11
                     /product="6-phosphogluconate dehydrogenase"
                     /protein_id="BAA11267.1"
                     /db_xref="GI:1064781"
                     /translation="GVFHTNWSE"
     gene            496. .1059
                     /gene="ahpC"
     CDS             496. .1059
                     /gene="ahpC"
                     /note="induced by hydrogen peroxide"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="alkyl hydroperoxide reductase"
                     /protein_id="BAA11268.1"
                     /db_xref="GI:1064782"
                     /translation="MSLIGKEVLPFEAKAFKNGEFIDVTNEDLKGQWSVFCFYPADFS
                     FVCPTELEDLQEQYAALKELGVEVYSVSTDTHFVHKGWHDSSEKISKITYAMIGDPSQ
                     TISRNFDVLDEETGLADRGTFIIDPDGVIQTVEINAGGIGRDASNLVNKVKAAQYVRQ
                     NPGEVCPAKWEEGGETLTPSLDLVGKI"
     gene            1073. .2602
                     /gene="ahpF"
     CDS             1073. .2602
                     /gene="ahpF"
                     /note="induced by hydrogen peroxide"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="alkyl hydroperoxide reductase"
                     /protein_id="BAA11269.1"
                     /db_xref="GI:1064783"
                     /translation="MVLDANIKAQLNQYMQLIENDIVLKVSAGEDDTSKDMLALVDEL
                     ASMSSKISVEKAELNRTPSFSVNRVGEDTGVTFAGIPLGHEFTSLVLALLQVSGRPPK
                     VDQKVIDQVKKISGEYHFESYISLTCHNCPDVVQALNMMSVLNPNITHTMIDGAAYKA
                     EVESKNIMAVPTVYLNGESFGSGRMTLEEILAKMGSGTDASEFADKEPFDVLVVGGGP
                     AGASAAIYTARKGIRTGVVAERFGGQVLDTMSIENFISVKATEGPKLAASLEEHVKEY
                     DIDVMNLQRAKRLEKKDLFELELENGAVLKSKTVILSTGARWRNVNVPGEQEFKNKGV
                     AYCPHCDGPLFEGKDVAVIGGGNSGIEAAIDLAGIVNHVTVLEFAPELKADEVLQKRL
                     YSLPNVTVVKNAQTKEITGDQSVNGITYVDRETGEEKHVELQGVFVQIGLVPNTEWLE
                     GTVERNRMGEIIVDKHGATSVPGLFAAGDCTDSAYNQIIISMGSGATAALGAFDYLIR
                     N"
     gene            complement(2712. .4151)
                     /gene="bglA"
     CDS             complement(2712. .4151)
                     /gene="bglA"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="phospho-bete-glucosidase"
                     /protein_id="BAA11270.1"
                     /db_xref="GI:1064784"
                     /translation="MGNMPKDFLWGGALAAHQFEGGWNQGGKGPSVVDVMTAGAHGVP
                     RKITDTIEENEFYPNHEAIDFYHRYKEDIALFAEMGLKCLRTSIGWSRIFPKGDEAEP
                     NEAGLQFYDDVFDELLKHGIEPVITLSHFEMPLHLAREYGGFRNRKVVDFFVNFAEAC
                     FTRYKDKVKYWMTFNEINNQMDVNNPLFLWTNSGVVVGENENAKEVMYQTAHHELVAS
                     ALAVAKGKDINPEFQIGAMVSHVPIYPFSSNPEDVMLAEEEMRQRYFFPDVQVRGYYP
                     SYALKEFEREGYNITFEDGDDEILRNGTVDYLGFSYYMSTTVKSDVKNDNTGDIVNGG
                     LPNGVENPYITSSDWGWAIDPTGLRYTLNRFYDRYQIPLFIVENGFGAVDTLEEDGKV
                     HDPERIQYLKSHIEALKKAVTYDGVDLIGYTPWGIIDIVSFTTGEMKKRYGMIYVDRD
                     NEGNGSMKRYKKDSFEWYKNVIQTNGEEL"
     gene            4733. .5449
                     /gene="yydK"
     CDS             4733. .5449
                     /gene="yydK"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11271.1"
                     /db_xref="GI:1064785"
                     /translation="MRMLKYQQIATEIETYIEEHQLQQGDKLPVLETLMAQFEVSKST
                     ITKSLELLEQKGAIFQVRGSGIFVRKHKRKGYISLLSNQGFKKDLEDFNVTSKVIELD
                     VRKPTPEAAENLNIGMDEDIYYVKRVRYINGQTLCYEESYYTKSIVTYLNNEIVSHSI
                     FHYIREGLGLKIGFSDLFLHVGQLNEEEAEYLGLEAGLPKLYIESIFHLTNGQPFDYS
                     KISYNYEQSQFVVQANSFLL"
     gene            complement(5766. .6488)
                     /gene="yydJ"
     CDS             complement(5766. .6488)
```

```
                     /gene="yydJ"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11272.1"
                     /db_xref="GI:1064786"
                     /translation="MKLEFKKSISNKVIIILGAMFVFLFLLGYFLLVGIDKVSNVTPE
                     MFFFSSYTVATQFGLMLFSFVIAFFINREYSNKNILFYKLIGENIYTFFYKKIAVLFL
                     ECFAFITLGLLIISLMYHDFSHFALLLFLFSAVILQYILIIGTISVLCPNILISIGVS
                     IVYWMTSVILVAISNKTFGFIAPFEAGNTMYPRIERVLQSDNMTLGSNDVLFIILYLV
                     SIIIINAIVLRFSKTRWIKMGL"
     gene            complement(6509. .7138)
                     /gene="yydI"
     CDS             complement(6509. .7138)
                     /gene="yydI"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11273.1"
                     /db_xref="GI:1064787"
                     /translation="MNIANYTLKVKGKTLLQDTDLHFSSGKINHVVGKNGVGKSQLAK
                     DFLLNNSKRIGRDIRQNVSLISSSSNIPNDVSKDFLLHFLSKKFDAKMIDKIAYLLNL
                     DNIDGKVLIKNLSDGQKQKLKLLSFLLEDKNIIVLDEITNSLDKKTVIEIHGFLNKYI
                     QENPEKIIINITHDLSDLKAIEGDYYIFNHQEIQQYHSVDKLIEVYINE"
     gene            complement(7288. .8046)
                     /gene="yydH"
     CDS             complement(7288. .8046)
                     /gene="yydH"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11274.1"
                     /db_xref="GI:1064788"
                     /translation="MKILKYEDGKYEVLVQNNVFIKDKKSGEYYKNSLNSLSDKQLLR
                     FKMYKEKVSPKFFYLFLSFTALMFILNYIHLIKLQNGLSSVFYGWKMWIIIVIYFIMN
                     IVLHELGHIYSLKFFGKNFDKVGFKLNFYVFPAFYVQLNETYMLSRNEKIIVHLFGLF
                     INYLLINTLELINQFTFSSEALTMAFMLFSSTLLWNLIPILNSDGYKILLAFLSLDEY
                     SRFKTNHWLVLTIQIIGIGLAVNSVVHWILYIVN"
     gene            complement(8027. .8986)
                     /gene="yydG"
     CDS             complement(8027. .8986)
                     /gene="yydG"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA11275.1"
                     /db_xref="GI:1064789"
                     /translation="MYNKTVSINLDSRCNASCDHCCFSSSPTSTTRMEKEYIRELVTE
                     FAKNKTIQVISFTGGEVFLDYKFLKELMEIIKPYEKQITLISNGFWGLSKKKVQEYFH
                     DMNSLNVIALTISYDEYHAPFVKSSSIKNILEHSRKYPDIDISLNMAVTKDKMSNHIL
                     EELGDSILGVKITKFPMISVGAAKTRIKQENIHKFYSLEDEDSLHCPGYDIVYHHDGE
                     IYPCCSPAIFETKITLREEYNQSFERTVEKLNSNLLLFILRKEGFKWFLNILKENNKI
                     EEFDIPYEFSSICGVCGSLFNSAEKINYFYPYMEKYYNENFKV"
     gene            complement(9044. .9193)

  Query Match             3.3%;  Score 44.6;  DB 1;  Length 36200;
  Best Local Similarity   65.0%;  Pred. No. 3.6;
  Matches   65;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     77 ttaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgccccctngatgaaagtca 136
          || |  |  ||| | ||  |   |    || || ||| | ||     |   |||| || |
Db   4337 TTGCGCACATCGGGCTGGATACGGTGACATTTGACGGCACGCCTTTTTCCTTGAAGGTAA 4278

Qy    137 aggaaggcgatgccgtcaaaaaaggccaagtgctgtttga 176
          | |||||||||| |||||||||||||| |||||||  ||||
Db   4277 AAGAAGGCGATACCGTCAAAAAAGGCGAAGTGCTTGTTGA 4238


RESULT  22
BSUB0021/c
LOCUS       BSUB0021              215534 bp    DNA     linear   BCT 26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 21 of 21): from 3999281
            to 4214814.
ACCESSION   Z99124 AL009126
VERSION     Z99124.1  GI:2636442
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 215534)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 215534)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1. .215534
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            complement(11. .820)
                     /gene="yxjF"
     terminator      complement(11. .37)
                     /gene="yxjF"
     CDS             complement(47. .820)
                     /gene="yxjF"
                     /function="unknown"
                     /note="similar to gluconate 5-dehydrogenase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15933.1"
                     /db_xref="GI:2636443"
                     /db_xref="SWISS-PROT:P42317"
                     /translation="MRKQVALVTGAAGGIRFEIAREFAREGASVIVSDLRPEACEKAA
                     SKLAEEGFDAAAIPYDVTKEAQVADTVNVIQKQYGRLDILVNNAGIQHVAPIEEFPTD
                     TFEQLIKVMLTAPFIAMKHVFPIMKKQQFGRIINIASVNGLVGFAGKSAYNSAKHGVI
                     GLTKVGALEGAPHGITVNALCPGYVDTQLVRNQLSDLSKTRNVPYDSVLEQVIFPLVP
                     QKRLLSVKEIADYAVFLASEKAKGVTGQAVVLDGGYTAQ"
     gene            complement(837. .1487)
                     /gene="yxjE"
     CDS             complement(837. .1487)
                     /gene="yxjE"
```

```
                     /function="unknown"
                     /note="similar to 3-oxoadipate CoA-transferase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15934.1"
                     /db_xref="GI:2636444"
                     /db_xref="SWISS-PROT:P42316"
                     /translation="MKEARKRMVKRAVQEIKDGMNVNLGIGMPTLVANEIPDGVHVML
                     QSENGLLGIGPYPLEGTEDADLINAGKETITEVTGASYFDSAESFAMIRGGHIDLAIL
                     GGMEVSEQGDLANWMIPGKMVKGMGGAMDLVNGAKRIVVIMEHVNKHGESKVKKTCSL
                     PLTGQKVVHRLITDLAVFDFVNGRMTLTELQDGVTIEEVYEKTEADFAVSQSVLNS"
gene                 complement(1484. .2200)
                     /gene="yxjD"
CDS                  complement(1484. .2200)
                     /gene="yxjD"
                     /function="unknown"
                     /note="similar to 3-oxoadipate CoA-transferase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15935.1"
                     /db_xref="GI:2636445"
                     /db_xref="SWISS-PROT:P42315"
                     /translation="MGKVLSSSKEAAKLIHDGDTLIAGGFGLCGIPEQLILSIRDQGV
                     KDLTVVSNNCGVDDWGLGLLLANKQIKKMIASYVGENKIFERQFLSGELEVELVPQGT
                     LAERIRAGGAGIPGFYTATGVGTSIAEGKEHKTFGGRTYVLERGITGDVAIVKAWKAD
                     TMGNLIFRKTARNFNPIAAMAGKITIAEAEEIVEAGELDPDHIHTPGIYVQHVVLGAS
                     QEKRIEKRTVQQASGKGEAK"
gene                 complement(2279. .3643)
                     /gene="yxjC"
CDS                  complement(2279. .3643)
                     /gene="yxjC"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15936.1"
                     /db_xref="GI:2636446"
                     /db_xref="SWISS-PROT:P42314"
                     /translation="MELIIILLALGLLMFTAYRGFSVILFAPICALFAVLLTDPSHVL
                     PFFSSIFMEKMAGFIKLYFPVFLLGAIFGKVVEMAGLAASIAKTIVRLVGAKRAILAI
                     VLMGAVLTYSGVSLFVVVFAVYPFAKNMFQEANIPKRLIPGTIALGAFTFTMDALPGT
                     PQIQNVIPTSFFKTDIYAAPWLGLMGAVIVLAAGMLYLESRRKKAQASGEGYGGFDSQ
                     NAPAPESIESAAEPDKSPIRHALAFVPLILVGAVNKYFTIYLPKWYPNGFNFPSIGLK
                     EFGRLDISSAAAIWSVEIALVIGIITTILFDWRSVFAQLKEGLNEGIGGALLASMNTG
                     AEYGFGGIIAALPGFHKLSSGISHTFTDPLVNGAVTTALAGITGSASGGMGIALSAM
                     SEQYLQAIQAYNIPPEVMHRVISMASGGMDTLPHNGAVITLSGRDGFDPPAILSRYFC
                     DHAH"
gene                 complement(3762. .4645)
                     /gene="yxjB"
terminator           complement(3762. .3788)
                     /gene="yxjB"
CDS                  complement(3797. .4645)
                     /gene="yxjB"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15937.1"
                     /db_xref="GI:2636447"
                     /db_xref="SWISS-PROT:P42313"
                     /translation="MKRTVDFSMFRCPLCDSSMDAASGKSLICTERGHTFDLSRHGYV
                     NFLTKPVKTSYGAELFEARSRLIGECGFFDPLHDAIAELISHPKSGHEAFTILDSGCG
                     EGSHLNALCGFDYAGKAAIGTGIDLSKDGILKASKAFKDLMWAVADVARAPFHDRQFD
                     VVLSIFSPSNYAEFHRLLKNDGMLIKVVPRSDYLIELRQFLYTDSPRRTYSNTAAVER
                     FTANAAHSRPVRLRYVKTLDQQAIHWLLKMTPLAWSAPKDRVSLLKEMKSADITVDVD
                     ILIGMK"
gene                 5261. .6489
                     /gene="yxjA"
CDS                  5261. .6454
                     /gene="yxjA"
                     /function="unknown"
                     /note="similar to pyrimidine nucleoside transport"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15938.1"
                     /db_xref="GI:2636448"
                     /db_xref="SWISS-PROT:P42312"
                     /translation="MYFLLNLVGLIVIMAVVFLCSPQKKKIKWRPIITLIVLELLITW
                     FMLGTKVGSWAIGKIGDFFTWLIACASDGIAFAFPSVMANETVDFFFSALLPIIFIVT
                     FFDILTYFGILPWLIDKIGWVISKASRLPKLESFFSIQMMFLGNTEALAVIRQQLTVL
                     SNNRLLTFGLMSMSSISGSIIGSYLSMVPATYVFTAIPLNCLNALIIANLLNPVHVPE
                     DEDIIYTPPKEEKKDFFSTISNSMLVGMNMVIVILAMVIGYVALTSAVNGILGVFVHG
                     LTNQTIFAYLFSPFAFLLGLPVHDAMYVAQLMGMKLATNEFVAMLDLKNNLKSLPPHT
                     VAVATTFLTSFANFSTVGMIYGTYNSILDGEKSTVIGRNVWKLLVSGIAVSLLSAAIV
                     GLFVW"
terminator           6462. .6489
                     /gene="yxjA"
gene                 complement(6925. .7167)
                     /gene="yxiT"
CDS                  complement(6925. .7167)
                     /gene="yxiT"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15939.1"
                     /db_xref="GI:2636449"
                     /db_xref="SWISS-PROT:P42311"
                     /translation="MKWNNMLKAAGIAVLLFSVFAYAAPSLKAVQAKTPTVSTHTYKK
                     IKELTYPQVHHVGNAAFEKKINQELKAYMDNRIRNT"
terminator           complement(7290. .7306)
                     /gene="yxiS"
gene                 complement(7290. .7602)
                     /gene="yxiS"
CDS                  complement(7312. .7602)
                     /gene="yxiS"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15940.1"
                     /db_xref="GI:2636450"
                     /db_xref="SWISS-PROT:P42310"
                     /translation="MNRKQLTEAYERDENMMILVFAQWCVNHDLDPMELYAKAYPQQK
                     LNESLKKTMDDLVVPKHEAEHIPDQTVIAVLEMFGNTDLAQALHEAISGRKQ"
gene                 complement(7652. .9712)
                     /gene="katB"
CDS                  complement(7652. .9712)
                     /gene="katB"
                     /EC_number="1.11.1.6"
                     /note="alternate gene name: katE"
                     /codon_start=1
```

```
  Query Match              3.3%;  Score 44.6;  DB 1;  Length 215534;
  Best Local Similarity   65.0%;  Pred. No. 3.7;
  Matches   65;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     77 ttaccgaagtcgcgttgcttggcgaagaatatgccggtatgcgcccctngatgaaagtca 136
          || |  |  ||| | ||  |   |    || || ||| | ||     |   |||| || |
Db 122719 TTGCGCACATCGGGCTGGATACGGTGACATTTGACGGCACGCCTTTTTCCTTGAAGGTAA 122660

Qy    137 aggaaggcgatgccgtcaaaaaaggccaagtgctgtttga 176
          | ||||||||| |||||||||||||| |||||||  ||||
Db 122659 AAGAAGGCGATACCGTCAAAAAAGGCGAAGTGCTTGTTGA 122620


RESULT  23
AB015974/c
LOCUS       AB015974               1641 bp    DNA     linear   BCT 28-AUG-1999
DEFINITION  Pseudomonas tolaasii glpK gene for glycerol kinase, complete cds.
ACCESSION   AB015974
VERSION     AB015974.1  GI:3358090
KEYWORDS    glpK; glycerol kinase.
SOURCE      Pseudomonas tolaasii (strain:PT814) DNA, clone_lib:pHHM105.
  ORGANISM  Pseudomonas tolaasii
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (sites)
```

```
  AUTHORS   Murata,H.
  TITLE     Molecular cloning of glpFKRD of a mushroom pathogenic bacterium, P
            seudomonas tolaasii strain PT814: induction of an avirulent mutant
            carrying single mini-Tn5km1 insertion in glpD
  JOURNAL   Mycoscience 40, 351-356 (1999)
REFERENCE   2  (bases 1 to 1641)
  AUTHORS   Murata,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1998) Hitoshi Murata, Forestry and Forest
            Products Research Institute, Bio-Resource Development;
            Matsunosato-1, Kukizaki-machi, Ibaraki 305-8687, Japan
            (E-mail:murmur@ffpri.affrc.go.jp, Tel:81-298-73-3211(ex.458),
            Fax:81-298-73-3795)
COMMENT     On Jul 30, 1998 this sequence version replaced gi:3299803.
FEATURES             Location/Qualifiers
     source          1. .1641
                     /organism="Pseudomonas tolaasii"
                     /strain="PT814"
                     /db_xref="taxon:29442"
                     /clone_lib="pHHM105"
     RBS             71. .74
     gene            83. .1594
                     /gene="glpK"
     CDS             83. .1594
                     /gene="glpK"
                     /codon_start=1
                     /transl_table=11
                     /product="glycerol kinase"
                     /protein_id="BAA31995.1"
                     /db_xref="GI:3358091"
                     /translation="MTDIQNKNYIIALDQGTTSSRAIIFDRDANVVCTAQREFTQHYP
                     QAGWVEHDPMEIFATQSAVMVEALAQAGLHHDQVAAIGITNQRETTVVWDKVTGRPIY
                     NAIVWQCRRSTEIQICQQLKRDGHEQYINDTTGLVTDPYFSGTKLKWILDNVEGSRER
                     ARNGELLFGTIDSWLIWKFTGGKTHVTDYTNASRTMLFNIHTLEWDAKMLEILDVPRE
                     MLPEVKSSSEIYGRTKVASPSAAIAGDQQAALFGQMCVRGGQAKNTYGTGCFLLMNTG
                     DKAVKSKHGMLTTIACGPRGEVAYALEGAVFNGGSTVQWLRDELKIIADATDTEYFAG
                     KVKDSNGVYLVPAFTGLGAPYWDPYARGALFGLTRGVRVDHIIRAALESIAYQTRDVL
                     DAMQQDSGERLKALRVDGGAVANNFLMQFQADILGTQVERPQMRETTALGAAYLAGLA
                     CGFWGSLEELRGKAVIEREFEPQLDEAAKEKLYAGWQKAVSRTRDWEPHEGAE"
BASE COUNT      346 a    541 c    492 g    262 t
ORIGIN
```

```
  Query Match             3.1%;  Score 41.2;  DB 1;  Length 1641;
  Best Local Similarity   47.6%;  Pred. No. 23;
  Matches  121;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy   1007 tcggctgggttgcgccgcagccggacaaatactccatcacgcgtacgaccctcggccatt 1066
          ||||    |  ||| | | ||| |||||   |   ||    ||  || ||   | || |
Db    423 TCGGTGCTGCGGCGGCACTGCCAGACAATGGCGTTATAGATCGGGCGGCCGGTGACCTTG 364

Qy   1067 tcctgaaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatgg 1126
          |||   |  ||      |||| | |       |  | | | || |||||||    |||||
Db    363 TCCCACACCACCGTGGTTTCACGCTGGTTGGTGATGCCGATGGCGGCGACCTGATCATGG 304

Qy   1127 tgccgattggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgc 1186
          ||| |   ||      | ||||| |     |  |    |   |||  |  |     ||| |
Db    303 TGCAGGCCGGCCTGGGCCAGCGCTTCGACCATCACCGCGCTCTGGGTGGCGAAAATTTCC 244

Qy   1187 gcgatttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacg 1246
            | ||  ||| |   |||| |     |  ||   || ||   ||    ||| ||
Db    243 ATCGGGTCATGCTCGACCCAACCGGCTTGCGGGTAGTGCTGAGTGAATTCACGCTGGGCG 184

Qy   1247 aagaagacctcgct 1260
            | ||||| || |
Db    183 GTGCAGACCACGTT 170
```

```
RESULT  24
AF316898
LOCUS       AF316898                 628 bp    DNA     linear   PLN 01-NOV-2001
DEFINITION  Haberlea rhodopensis internal transcribed spacer 1, 5.8S ribosomal
            RNA gene, and internal transcribed spacer 2, complete sequence.
ACCESSION   AF316898
VERSION     AF316898.1  GI:16565500
KEYWORDS    .
SOURCE      Haberlea rhodopensis.
  ORGANISM  Haberlea rhodopensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Lamiales; Gesneriaceae; Haberlea.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Moeller,M. and Cronk,Q.C.B.
  TITLE     Evolution of morphological novelty: A phylogenetic analysis of
            growth patterns in Streptocarpus (Gesneriaceae)
  JOURNAL   Evolution (2001) In press
REFERENCE   2  (bases 1 to 628)
  AUTHORS   Moeller,M. and Cronk,Q.C.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2000) Systematics and Evolution, Royal Botanic
            Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Haberlea rhodopensis"
                     /isolate="E1"
                     /specimen_voucher="19754106"
                     /db_xref="taxon:121482"
     misc_RNA        1. .233
                     /product="internal transcribed spacer 1"
     rRNA            234. .397
                     /product="5.8S ribosomal RNA"
     misc_RNA        398. .628
                     /product="internal transcribed spacer 2"
BASE COUNT      145 a    177 c    182 g    124 t
ORIGIN
```

```
  Query Match             3.1%;  Score 41;  DB 8;  Length 628;
  Best Local Similarity   47.9%;  Pred. No. 26;
  Matches  104;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    144 cgatgccgtcaaaaaaggccaagtgctgtttgaagacaaaaagnatccgggcgtggtgtt 203
          |||  ||  ||||  || ||        | |  |   |||||   |  || ||||| ||
Db      2 CGAAACCTGCAAAGCAGACCCGCGAACATGTTTAATGAAAAATATTGTGGCCGTGGCGTC 61

Qy    204 taccgcgccngtttcaggcaaaatcgccgccatccatcgcggcgaaaagcgcgtacttca 263
            ||  | ||  | ||||  |  ||| |  |||   || |||  |||    |
Db     62 GGATGCTTCGGTGCCCGGCATCACGGCCCCGGTCCCAAGCAGCGCCAAGTTGGCTTGGGC 121

Qy    264 gtcggtcgtgattgccgttgaaggcaacgacgaaatcgagttcgaacgctacgcgcccga 323
          ||         | | ||  | | ||  | | |||| | |  |||||| |  | |   |
Db    122 GTACTAACCTCTCGGCGCGGCAAGCGCCAAGGAAAACCATATCGAACACCTCTCCGTCCG 181

Qy    324 agcgttggcaaacttaagcggcgangaantnngnngc 360
           |||| | |||||     |   || |   |  | ||
Db    182 GGCGTCGTCAAACGGTGCCAAGGATGTGATGAGGAGC 218
```

```
RESULT  25
AE002001/c
LOCUS       AE002001                 11375 bp    DNA     linear   BCT 22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 138 of 229 of the complete
            chromosome 1.
ACCESSION   AE002001 AE000513
VERSION     AE002001.1  GI:6459336
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcaceae;
            Deinococcus.
REFERENCE   1  (bases 1 to 11375)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.
```

```
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 11375)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11375
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            complement(101. .1885)
                     /gene="DR1571"
     CDS             complement(101. .1885)
                     /gene="DR1571"
                     /note="similar to GB:AE000657 percent identity: 51.51;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="peptide ABC transporter, periplasmic
                     peptide-binding protein"
                     /protein_id="AAF11135.1"
                     /db_xref="GI:6459337"
                     /translation="MKKVMMLALALGASTSLAAPFVYPANWTSNKPGDVQTGGTFRSV
                     NLQDFKTLNPFVSSESPNLPAVLSAGSLLGYNPVTGNYAPYMAEKYTQSADKRTFTFD
                     IRKGMKWSDGKPITVDDWITAYTIDSNKDVGSNTFDYWTINNQPIKVTKVDSNTLKVV
                     FPKADVTAIEFLSGIFLPQPTHVFMPVWKAKGAQGIKDMWTISTNPDNIVTSGPFMLD
                     RYVRGERAILKKNPYFGEWNKDSAGKSLPYLDGIQINIVADANAQLAQFLAGNLDTYS
                     PDNRDKLAQVKSAMDGGKVKGTLIPNASARASSDFMVFNMDDSATFKTKLFSNVKFRQ
                     AMSMLMNRDAMVDLALGGLGEPTYTSVYPVYKDWIPSGMDKYKFNPTAAAKLLAELGF
                     TKKGSDGILVDKAGNKLEFTLITNAENNRRQSYAKVIQDEAKKVGVKINVSAIAFNQM
                     TTLLDAKDNFGRRNFDAIIIGLTGGGQVYPVSGPSVVECKGLGDGGNLHMFNQSNKCR
                     FPFETQAVNLFWKGRAEFDLAKRKAIAAQIQRNEMENQPYIQLAAQTVHFAWTDRVQG
                     EYNRPQINSLNASTLFGPRDIALTWIKR"
     gene            2042. .4308
                     /gene="DR1572"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     GB:AL009126 percent identity: 48.68; identified by
                     sequence similarity; putative;helicase-related protein,
                     authentic frameshift"
     gene            4529. .6163
                     /gene="DR1573"
     CDS             4529. .6163
                     /gene="DR1573"
                     /note="similar to GB:M22039 SP:P13242 PID:143597
                     PID:853762 GB:AL009126 percent identity: 73.03; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="CTP synthase"
                     /protein_id="AAF11136.1"
                     /db_xref="GI:6459338"
                     /translation="MKYIFVTGGVVSSLGKGVASASLGALLRARGYKVTAVKIDPYIN
                     IDAGTMRPYEHGECFVTASGAETDLDIGNYERFLDLDIPAGSNITTGQVYQEVIRKER
                     AGDYLSQTVQVIPHVTDEIKRRIRAAGESAGAEIVLIEVGGTVGDIESLPFLEAIRQF
                     RFDEGDDNVLYLHLTLVPYLGTSNEFKTKPTQHSVAALRSYGISPDIVMVRSKDKLPA
                     EITRKIAAFTSVRENRVFSSYDVEHVYQVPLALEEQGLGKAVEELLGLEGTHPRLGVW
                     TDAVRTIKHPKQQVTIAIAGKYTAMPDAYLSLMESLTHAGIANDARVNIKWVNAEQLA
                     EAGAGELEGQFADVDGILVPGGFGIRGIEGKVKAAEYARTHGVPYLGICLGMQIAVIE
                     YARHVAGIEDANSAEFDEYAKNKVIDLMPEQLEVAGMGGTMRLGDWPMELRAGTKIAE
                     LYGVPQGGTVKERHRHRYEVNPAYVEQLEAAGLTISGVTPGVEGRGAGLVETVEIADH
                     PFFVALQAHPEFKSRPMRPSPPFAGFVKAALEHQQQ"
     gene            6206. .6922
                     /gene="DR1574"
     CDS             6206. .6922
                     /gene="DR1574"
                     /note="similar to GB:L77117 SP:Q58227 PID:1591506 percent
                     identity: 49.48; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF11137.1"
                     /db_xref="GI:6459339"
                     /translation="MISASGGNVGVRLARYAGPMRLPRRFPLLAALAAGAWYLRAVYR
                     FRDPVRLPPADQSGAGAVLSPADGLVSFVRRIEGGQVNGQNVAELLGTPQRDGWLLGL
                     YVGPLDVHFTYAPVGGTVISAGKRDGTRQGLPLAETARLLAGQPVDLLGSAAVRGNER
                     YSYAVQGEGGQTVTVGLIAPGAGLQAITYLDEGQTVRQGNKAAFLAEGGLVLLALPGS
                     VTPQVSVGERVRGAETVVAC"
     gene            complement(6927. .7574)
                     /gene="DR1575"
     CDS             complement(6927. .7574)
                     /gene="DR1575"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF11141.1"
                     /db_xref="GI:6459343"
                     /translation="MPFSELLTYFQPLAAEHYDLGPCAILQTPGANILALNAAFFPAE
                     AAENLPDVRRWFSARGRPALLLSDAALLGEDVGGLRVGSWTAGETRENIIVEQVSRLQ
                     LSLWSAVLAESYETLEWAPALARHFAARLEGDRASTLLLAYADGDPVGAGLWRTGPGG
                     TGAMHLWGTLDPAADAPLLNAAADLSGGNVRVSLPDTSPLSVPEAEVLTFTLLDG"
     gene            complement(7614. .7937)
                     /gene="DR1576"
     CDS             complement(7614. .7937)
                     /gene="DR1576"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF11142.1"
                     /db_xref="GI:6459344"
                     /translation="MMFRDPQPDLTPAHPLAPMQLSQAAQAYLEGCPAAQRDRAAALL
                     TALEDYLGAPAPLLAYTKLTGEAWQRTLPASEQTAAAALLADFRAFLRDGGWLDAARA
                     VNLPD"
     gene            complement(7984. .9072)
                     /gene="DR1577"
     CDS             complement(7984. .9072)
                     /gene="DR1577"
                     /note="similar to GB:M37702 SP:P21516 PID:147485 GB:U00096
                     PID:1773089 percent identity: 62.61; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="s-adenosylmethionine:tRNA
                     ribosyltransferase-isomerase"
                     /protein_id="AAF11138.1"
                     /db_xref="GI:6459340"
                     /translation="MPDSTADPSTADEVLARLHFDLPEDRIAQTGAEPRDTSRLMVVG
                     EQIKHRMFRDLPDLLRPSDLLVFNESRVIPARVLARKPVVNGFGGGQVEVLLLREEFD
                     VGANVWSAYLKPAKRAGHELWLGENEATGHRAEVVGVLDDGARLLRFDHDIKPHLDEI
                     GRLPLPPYINAGDSDETWRERYQTVYAKTPGSVAAPTAGLHFTPELLARLDEMGVERA
                     SVTLHVGAGTFKPIQGPVADHVMHAERYEVSETNAAAITRAKAEGRRVVAVGTTTVRT
                     LESAWDGAAVRAGAGETRIFITPGTRVEVPDLLITNLHLPGSTLLLLVAAFAGEDRIR
                     AAYDAALSQDYRFYSLGDAMLLERLTSA"
     gene            complement(9076. .9633)
                     /gene="DR1578"
     CDS             complement(9076. .9633)
                     /gene="DR1578"
                     /note="similar to PID:1001613 PID:1001641 percent
                     identity: 56.10; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="conserved hypothetical protein"
                     /protein_id="AAF11139.1"
                     /db_xref="GI:6459341"
                     /translation="MQYVVHRPRVGVFIDTQNLYHSARDLLERTVNFETILQVATEGR
                     ELVHAISYTVEREGEATSRPFIYKLSTLGFKVRRMNLTLHHVTDGGKPIYEGNWDMGI
                     VADMFRLMDHLDVIVLGSGDGDFTDIVEVLQERGKRVEVIAFREHTAQKLIDAADRFT
                     HLPDIEEGLMPARHKPGAKDAAGEA"
     gene            9827. .11185
                     /gene="DR1579"
     CDS             9827. .11185
                     /gene="DR1579"
                     /note="similar to GB:AL123456 percent identity: 62.91;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF11140.1"
                     /db_xref="GI:6459342"
                     /translation="MDPASGPGGWTKTSRAPAPARLRYPDPMTSSATPDALAQAQAAY
                     AELKARGLKLNLQRGQPADADFDLSNPMLSVLGENDTRYDGTDLRNYPGGVAGLPSAR
                     ALFGNYLDLKPENVLVWNNSSLELQGLVLGFALLHGLRGSQGGWVHEKPKMIVTVPGY
                     DRHFLLLETLGFKLLAVDMQDDGPDVDAIERLAANDASVKGILFVPTYSNPGGETISA
                     EKARRLVGLKAAAPDFTIFADDAYRAHHLFGEGERDVPVNFVALCRDAGHPDRAFVFA
                     STSKITFAGAGLGFVGSSEDNIGWLGKYLGAQSIGPNKFEQARHVKFLEGYDGGLEGL
                     MRDHAALIAPKFQAVYDALSAGLGEGGGGYATWRTPRGGYFISLDTVDPVAERVVQLA
                     DEAGISLTPAGATYPGGNDPKNANIRLAPTRPPLAEVEEAMRGVATCVKLAAEEYRAG
                     QH"
BASE COUNT     1983 a   3660 c   3871 g   1861 t
ORIGIN


  Query Match             3.0%;  Score 40.6;  DB 1;  Length 11375;
  Best Local Similarity   47.5%;  Pred. No. 33;
  Matches  115;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy      60 ttatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggtatgcg 119
           | |||    || ||||  |  ||||||||||  ||| ||||   |    | |||  |  |
Db    4087 TGATGGGCTGCGCGTCGGTGTCGAAGTCGCGCAGCTCGGCGCTCAGGCGGTCGGCGTCGG 4028

Qy     120 cccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttgaaga 179
           ||||  |  |     | | |  |  | ||  ||        ||||      |  |||| |
Db    4027 CCCCGCGCCGCGTCACGATGGCGATGTTGGTGTGCCCGGCGGCCTGCGCGTCCTTGACGG 3968

Qy     180 caaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatcca 239
           |       ||| | ||| | |  | |||||  |  ||  |      |  ||| | || ||
Db    3967 CCTGGGCAATCAGCGCGCGCTCGTCACCGCCCTCGTAGCGCTGCACCTCGGCCCCCTCGC 3908

Qy     240 tcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaat 299
             |  ||    || | | |      |||| ||   |  ||   |      || |||  |
Db    3907 GGTCCACGCCCTGCACCTGCGCGGCGCGGTTGTAGGTTTCGGCAATCCGGGCGCCGAGTT 3848

Qy     300 cg 301
           ||
Db    3847 CG 3846


RESULT  26
NMA1Z2491/c
LOCUS       NMA1Z2491              340806 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 1/7.
ACCESSION   AL162752 AL157959
VERSION     AL162752.2  GI:7378778
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 340806)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 340806)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .340806
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            complement(38. .358)
                     /gene="NMA0001"
     CDS             complement(38. .358)
                     /gene="NMA0001"
                     /note="NMA0001, unknown, len: 106 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0001"
                     /protein_id="CAB83321.1"
                     /db_xref="GI:7378779"
                     /db_xref="SPTREMBL:Q9JX93"
                     /translation="MSNTQTIRSADFTTSRAWGALDIANMNGTTVRLHWTDQPYKWHV
                     NDGEEVFAVMDGEVDMHYRENGEEHIVRLKSGDIFYAGIGTEHVAHPRGETRILVIEK
                     EGSV"
     gene            complement(414. .2438)
                     /gene="nuoL"
     CDS             complement(414. .2438)
                     /gene="nuoL"
                     /EC_number="1.6.5.3"
                     /note="NMA0002, nuoL, NADH dehydrogenase I chain L, len:
                     674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase
                     I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to
                     entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone
                     (complex I), various chains, and to entry PF00662
                     oxidored_q1_N, NADH-Ubiquinone oxidoreductase (complex I),
                     chain 5 N-terminus"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain L"
                     /protein_id="CAB83322.1"
                     /db_xref="GI:7378780"
                     /db_xref="SWISS-PROT:Q9JX92"
                     /translation="MNDMTLYLIIALVPLAGSLIAGLFGNKIGRAGAHTVTILGVAVS
                     AVLSAYVLWGFLNGSRAKFDENVYTWLTMGGLDFSVGFLVDTMTAMMMVVVTGVSLMV
                     HIYTIGYMHDEKVGYQRFFSYISLFTFSMLMLIMSNNFIQLFFGWEAVGLVSYLLIGF
                     YFKRPSATFANLKAFLINRVGDFGFLLGIGLVLAYFGGSLRYQDVFAYLPNVQNATIQ
                     LFPGVEWSLITVTCLLLFVGAMGKSAQFPLHVWLPDSMEGPTPISALIHAATMVTAGL
                     FMVSRMSPIYEMSSTALSVIMVIGAITALFMGFLGVIQNDIKRVVAYSTLSQLGYMTV
                     ALGASAYSVAMFHVMTHAFFKALLFLAAGSAIIGMHHDQDMRHMGNLKKYMPITWLTM
                     LIGNLSLIGTPFFSGFYSKDSIIEAAKYSTLPGSGFAYFAVLASVFVTAFYAFRQYFM
                     VFHGEEKWRSLPEHHSDGHGEEHHGLGKNDNPHESPLVVTLPLILLAVPSVIIGYIAI
                     EPMLYGDFFKDVIFVNADAHPTMHIMKEEFHGALAMVSHSLHSPVLYLAIAGVLSAWL
                     LYVKLPHLPAKIAQAFRPVYVLFENKYYLDALYFNVFAKGTRALGTFFWKVGDTAIID
                     NGIVNGSARLVGAVAAQVRKVQTGFIYTYAAAMVFGVLVLLGMTFWGLFR"
     misc_feature    1035. .1044
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(1116. .2030)
                     /gene="nuoL"
                     /note="Pfam match to entry PF00361 oxidored_q1,
                     NADH-Ubiquinone/plastoquinone (complex I), various chains,
```

```
                     score 351.80, E-value 7.3e-102"
     misc_feature    complement(2061. .2243)
                     /gene="nuoL"
                     /note="Pfam match to entry PF00662 oxidored_q1_N,
                     NADH-Ubiquinone oxidoreductase (complex I), chain 5
                     N-terminus, score 77.20, E-value 3.5e-19"
     gene            complement(2487. .2828)
                     /gene="NMA0003"
     CDS             complement(2487. .2828)
                     /gene="NMA0003"
                     /note="NMA0003, unknown, len: 113 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0003"
                     /protein_id="CAB83323.1"
                     /db_xref="GI:7378781"
                     /db_xref="SPTREMBL:Q9JX91"
                     /translation="MMDKNQLEQEFHKAMLNIYQEALNLPQPYKATRFLQIVNEFGGK
                     EAADKLLSTGEKKTQTGFTELILSGGGVHALKYSMEYLVLQKPWCDLFTEEQLAVARK
                     RLERVGFVSPK"
     RBS             complement(2833. .2837)
     gene            complement(2952. .3527)
                     /gene="NMA0004"
     CDS             complement(2952. .3527)
                     /gene="NMA0004"
                     /note="NMA0004, unknown, len: 191 aa; similar to
                     hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa),
                     fasta scores; E(): 0; 80.5% identity in 190 aa overlap,
                     and to proteins involved in cell cycle e.g. TR:O54679
                     (EMBL:AF036487) Lactococcus lactis Plasmid pNZ4000
                     putative mobilization protein (200 aa), fasta scores; E():
                     0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell
                     filamentation protein FIC (200 aa), fasta scores; E():
                     0.093, 29.4% identity in 126 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0004"
                     /protein_id="CAB83324.1"
                     /db_xref="GI:7378782"
                     /db_xref="SPTREMBL:Q9JQR9"
                     /translation="MPSENPIGKTMKSIDEQSLHNARRLFESGDIDRIEVGTTAGLQQ
                     IHRYLFGGLYDFAGQIREDNISKGGFRFANAMYLKEALVKIEQMPERTFEEIIAKYVE
                     MNIAHPFLEGNGRSTRIWLDLVLKKNLKKVVNWQNVSKTLYLQAMERSPVNDLELRFL
                     LKDNLTDDVDNREIIFKGIEQSYYYEGYEKG"
     misc_feature    complement(3516. .3525)
                     /gene="NMA0004"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(3561. .3866)
                     /gene="nuoK"
     CDS             complement(3561. .3866)
                     /gene="nuoK"
                     /EC_number="1.6.5.3"
                     /note="NMA0005, nuoK, NADH dehydrogenase I chain K, len:
                     101 aa; simlar to many e.g. NUOK_RHOCA NADH dehydrogenase
                     I chain K (EC 1.6.5.3) (102 aa), fasta scores; E():
                     1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam
                     match to entry PF00420 oxidored_q2,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain K"
                     /protein_id="CAB83325.1"
                     /db_xref="GI:7378783"
                     /db_xref="SPTREMBL:Q9JQU9"
                     /translation="MITLTHYLVLGALLFGISAMGIFMNRKNVLVLLMSIELMLLAVN
                     FNFIAFSQHLGDTAGQIFVFFVLTVAAAESAIGLAIMVLVYRNRQTINVADLDELKG"
     misc_feature    complement(3564. .3863)
                     /gene="nuoK"
                     /note="Pfam match to entry PF00420 oxidored_q2,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
                     score 141.50, E-value 1.5e-38"
     gene            complement(3863. .4534)
                     /gene="nuoJ"
     CDS             complement(3863. .4534)
                     /gene="nuoJ"
                     /EC_number="1.6.5.3"
                     /note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
                     223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
                     I chain J (EC 1.6.5.3) (202 aa), fatsa scores; E():
                     9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
                     match to entry PF00499 oxidored_q3,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain J"
                     /protein_id="CAB83326.1"
                     /db_xref="GI:7378784"
                     /db_xref="SPTREMBL:Q9JX90"
                     /translation="MTFQLILFYIFAVIILYGAIKTVTAKNPVHAALHLVLTFCVSAM
                     LWMLMQAEFLGVTLVVVYVGAVMVLFLFVVMMLNIDIEEMRAGFWRHAPVAGVVGTLL
                     AVALILILVNPKTDLAAFGLMKDIPADYNNIRDLGSRIYTDYLLPFELAAVLLLLGMV
                     AAIALVHRKTVNPKRMDPADQVKVRADQGRMRLVKMEAVKPQTESAEESEVSDDLKPK
                     EEGKA"
     RBS             complement(3874. .3879)
                     /gene="nuoJ"
     misc_feature    complement(4034. .4525)
                     /gene="nuoJ"
                     /note="Pfam match to entry PF00499 oxidored_q3,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
                     score 77.90, E-value 2e-19"
     gene            complement(4570. .5277)
                     /gene="NMA0007"
     CDS             complement(4570. .5277)
                     /gene="NMA0007"
                     /note="NMA0007, unknown, len: 235 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0007"
                     /protein_id="CAB83327.1"
                     /db_xref="GI:7378785"
                     /db_xref="SPTREMBL:Q9JX89"

  Query Match              2.9%;  Score 39.2;  DB 1;  Length 340806;
  Best Local Similarity   53.2%;  Pred. No. 77;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    870 cgaattggttgacgcagacaaccgcgtgatttccggttcggtattgaacggcgcgattac 929
          ||||||||||  ||    | | |||    |  ||      | | | |||| ||||||
Db 230044 CGAATTGGTTTCCGTCCTCGAACGCCCTGTAACCAAAGAAGAAATCAACGCCGCGATGAA 229985

Qy    930 acaaggcgcgcacgattatttgggacgctaccacaatcagatttccgttatcgaagaagg 989
          |   | |||   ||| | ||  ||   | || |  |||| |||     |  ||| |  |
Db 229984 AGCCGCCGCCAGCGAATCTTACGGCTACAACGAAGATCAAATTGTTTCTTCCGACGTTGT 229925

Qy    990 ccgcagcaaagagctgttcggctgggttgcgccgca 1025
          | ||| | || |    |    ||  | |||| | ||
Db 229924 CGGCATCGAATACGGTTCGCTCTTCGATGCGACCCA 229889


RESULT  27
AY071018
LOCUS       AY071018               2557 bp    mRNA    linear   INV 20-DEC-2001
DEFINITION  Drosophila melanogaster RE09574 full length cDNA.
ACCESSION   AY071018
VERSION     AY071018.1  GI:17945161
KEYWORDS    FLI_CDNA.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2557)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
```

```
            Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
            and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2001) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone. However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .2557
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /map="57B13-57B13"
     gene            1. .2557
                     /gene="CG15651"
                     /note="alignment with genomic scaffold AE003452"
                     /db_xref="FLYBASE:FBgn0034567"
     CDS             294. .1877
                     /gene="CG15651"
                     /note="Longest ORF"
                     /codon_start=1
                     /db_xref="FLYBASE:FBgn0034567"
                     /product="RE09574p"
                     /protein_id="AAL48640.1"
                     /db_xref="GI:17945162"
                     /translation="MKVFRLRLVKLLLLGFLLANLVFLYYTWNTLLWRRISRALSGES
                     QPAEEPPKAAKLSPKEKLRNANKHIRKSITIVFYGHYNFEQDLKASIDSILDVIPNMP
                     ILVLQDGGAEYAYPPVNYERNLTAGEERTVLFQSLSFDVRRTRQELNPLASIRTKYAL
                     IMPDGVRLNSKNILQKILREINSMGLPGQPKEQRPPVPPEQLVRRMVLVPFAGNLKSF
                     SSCVKMNLDLPDWTLELVATNDTRRCDLFLQKHAILLDAAALAAMPEPFSLPFPEMLY
                     MQAKIANLSTTVFPQAFQEGRRLFASFHTKQRRMDLRRRQFREMYKKLQIKRIVRRAY
                     RVPGKAQAREVWGQGHGLVLDGQFSSSNTSLPLITDIDLFGCERTTKSCIGSVYNERP
                     YYSYLGKHTPPCCLDKLRTTFNHVLEEFENVGIRYWLDNRALQSAIETNHLSPDAYDI
                     DISFNVQDLERSNAMKKSQSKPYVDNEGFYWIKATDGHHFRVQFSKPNQVGVNLLPYS
                     ISGTEAKASGFFGWKARSF"
BASE COUNT      699 a    629 c    631 g    598 t
ORIGIN


  Query Match             2.8%;  Score 38.2;  DB 3;  Length 2557;
  Best Local Similarity   46.4%;  Pred. No. 1.3e+02;
  Matches  121;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy   1072 aaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccg 1131
          || ||  |  |  ||||| ||  ||   |||   |  |||| ||  |  | | | || |
Db    489 AACAAACACATACGCAAGTCCATCACCATCGTGTTCTATGGCCACTACAACTTTGAGCAG 548

Qy   1132 attggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgat 1191
            |   |       |  ||  |     || ||| |  | || | | |||   | |  |
Db    549 GATCTGAAGGCATCGATCGACAGCATCCTGGACGTGATACCCAACATGCCCATCCTGGTG 608

Qy   1192 ttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaa 1251
            |    |  |||| | ||||   |||  |  |   ||    ||  |||  | ||  |
Db    609 CTTCAGGATGGCGGTGCCGAGTACGCCTATCCGCCGGTCAACTACGAACGGAACCTAACC 668

Qy   1252 gacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgt 1311
          | |  ||    | ||| | |  | | || | |    | |    | |  | ||   |||
Db    669 GCCGGCGAGGAGCGCACCGTGCTATTCCAGAGTCTGTCCTTCGATGTGCGGCGAACGCGG 728

Qy   1312 aaggtgctggaaaccnttgag 1332
           ||| |||| |  || | | |
Db    729 CAGGAGCTGAATCCCTTGGCG 749


RESULT  28
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a   25171 c   22273 g    9426 t
ORIGIN


  Query Match             2.8%;  Score 38.2;  DB 6;  Length 65140;
  Best Local Similarity   45.6%;  Pred. No. 1.3e+02;
  Matches  118;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    193 ggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccatcgcggcgaaaag 252
          |||| || | |   ||||   ||   | ||     || | ||  | |   || |    |
Db   5971 GGCGCGGAGATCTTCGCGGACGTGGAATGCCCGCCCGGCACCGCCGACGACGCCCCCGAC 6030

Qy    253 cgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatcgagttcgaacgc 312
          | || ||| ||  | | | || | | ||  |   |  ||| ||  || | | | ||  ||
Db   6031 CACGGACTGCACCCCGCCCTGCTCGACGCGGCCCGGCACGCCGCCATGGCGGTGGACGGC 6090

Qy    313 tacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngcaatctgatccaa 372
            || |||||  || | |     | |  |     | |   |  |   ||    | | |
Db   6091 ACCGTGCCCGTCGCCTGGCACGGCGTCCGGCTGCACGCCGTCGGCGCCACCGCGCTGCGG 6150

Qy    373 tccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccctgccgtcgatgcc 432
          | |  |  | | |  |  |     |   | | |  | ||        | |||  | | |
Db   6151 GTCCGCATCCGCCCCACCACGACCGGCACGCTGACCCTCACCGCGGTCGACGTGCACGGC 6210

Qy    433 gagccgttcgccatcttcg 451
          | |||| ||| || | |||
Db   6211 GCGCCGGTCGTCACCGTCG 6229


RESULT  29
AC005268
LOCUS       AC005268               89791 bp    DNA     linear   INV 07-JUL-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1s DS03550 (D225) and
            DS02397 (D206)), complete sequence.
```

```
ACCESSION   AC005268 AC004330 AC004314 AC003599
VERSION     AC005268.1  GI:3293207
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in sac from P1 clones DS03550
            (D225) and DS02397 (D206)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 89791)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome, region 57B6-57C1
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 89791)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov.
            Library locations: 137-37, 136-25.
FEATURES             Location/Qualifiers
     source          1. .89791
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="57B6-57C1"
                     /clone="P1s DS03550 (D225) and DS02397 (D206)"
                     /note="DS03550 (d225) extends from bp 1 to bp 16,924 and
                     DS02397 (D206) extends from bp 13,480 to bp 89,791."
BASE COUNT    24290 a  20016 c  20807 g  24678 t
ORIGIN
```

```
  Query Match             2.8%;  Score 38.2;  DB 3;  Length 89791;
  Best Local Similarity   46.4%;  Pred. No. 1.3e+02;
  Matches  121;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy   1072 aaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccg 1131
          || ||  |  |   ||||| ||  ||   |||   |  |||| ||  |  | | | || |
Db  78595 AACAAACACATACGCAAGTCCATCACCATCGTGTTCTATGGCCACTACAACTTTGAGCAG 78654

Qy   1132 attggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgat 1191
            |   |       |  ||  |     || ||| |  | || | | |||   | | |
Db  78655 GATCTGAAGGCATCGATCGACAGCATCCTGGACGTGATACCCAACATGCCCATCCTGGTG 78714

Qy   1192 ttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaa 1251
           |   |  |||| | ||||   |||  |  |   ||    ||  |||  | ||  |
Db  78715 CTTCAGGATGGCGGTGCCGAGTACGCCTATCCGCCGGTCAACTACGAACGGAACCTAACC 78774

Qy   1252 gacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgt 1311
          | |  ||    | ||| | |  | | || | |    | |    | |  | ||    |||
Db  78775 GCCGGCGAGGAGCGCACCGTGCTATTCCAGAGTCTGTCCTTCGATGTGCGGCGAACGCGG 78834

Qy   1312 aaggtgctggaaaccnttgag 1332
           ||| |||| |  || | | |
Db  78835 CAGGAGCTGAATCCCTTGGCG 78855
```

```
RESULT  30
AC018341/c
LOCUS       AC018341              101085 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC018341
VERSION     AC018341.1  GI:6552850
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 101085)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214413 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .101085
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    28299 a  22687 c  22278 g  27821 t
ORIGIN
```

```
  Query Match             2.8%;  Score 38.2;  DB 2;  Length 101085;
  Best Local Similarity   46.4%;  Pred. No. 1.3e+02;
  Matches  121;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy   1072 aaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccg 1131
          || ||  |  |   ||||| ||  ||   |||   |  |||| ||  |  | | | || |
Db  92582 AACAAACACATACGCAAGTCCATCACCATCGTGTTCTATGGCCACTACAACTTTGAGCAG 92523

Qy   1132 attggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgat 1191
            |   |       |  ||  |     || ||| |  | || | | |||   | | |
Db  92522 GATCTGAAGGCATCGATCGACAGCATCCTGGACGTGATACCCAACATGCCCATCCTGGTG 92463

Qy   1192 ttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaa 1251
           |   |  |||| | ||||   |||  |  |   ||    ||  |||  | ||  |
Db  92462 CTTCAGGATGGCGGTGCCGAGTACGCCTATCCGCCGGTCAACTACGAACGGAACCTAACC 92403

Qy   1252 gacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgt 1311
          | |  ||    | ||| | |  | | || | |    | |    | |  | ||    |||
Db  92402 GCCGGCGAGGAGCGCACCGTGCTATTCCAGAGTCTGTCCTTCGATGTGCGGCGAACGCGG 92343

Qy   1312 aaggtgctggaaaccnttgag 1332
           ||| |||| |  || | | |
Db  92342 CAGGAGCTGAATCCCTTGGCG 92322
```

```
RESULT  31
AF263912
LOCUS       AF263912              123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
```

```
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
     gene            4714. .5748
                     /gene="nysDIII"
     CDS             4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
     gene            5930. .34363
                     /gene="nysI"
     CDS             5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                     AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                     HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                     ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                     DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                     ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                     DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                     VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                     GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                     AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                     LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                     LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                     QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                     LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
```

```
DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGVRFADAVADLLAAEYRAFV
EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
TAGRPAGTWTPRGTTLVTGGSGTLAPHLARWLAQRGAEHLVLISRRGTAAPGAAELVA
```

```
Query Match              2.8%;  Score 38.2;  DB 1;  Length 123580;
Best Local Similarity   45.6%;  Pred. No. 1.3e+02;
Matches  118;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    193 ggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccatcgcggcgaaaag 252
          |||| || | |   ||||   ||   | ||     || | ||  | |   || |    |
Db  65824 GGCGCGGAGATCTTCGCGGACGTGGAATGCCCGCCCGGCACCGCCGACGACGCCCCCGAC 65883

Qy    253 cgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatcgagttcgaacgc 312
          | || ||| ||  | | | || | | ||  |   |  ||| ||  || | | | ||  ||
Db  65884 CACGGACTGCACCCCGCCCTGCTCGACGCGGCCCGGCACGCCGCCATGGCGGTGGACGGC 65943

Qy    313 tacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngcaatctgatccaa 372
            || |||||  || | |     | |  |     | |   |  |   ||    | | |
Db  65944 ACCGTGCCCGTCGCCTGGCACGGCGTCCGGCTGCACGCCGTCGGCGCCACCGCGCTGCGG 66003

Qy    373 tccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccctgccgtcgatgcc 432
          | |  |  |  |  |   |     |   | | |  ||        | |||  | | |
Db  66004 GTCCGCATCCGCCCCACCACGACCGGCACGCTGACCCTCACCGCGGTCGACGTGCACGGC 66063

Qy    433 gagccgttcgccatcttcg 451
          | |||| ||| || | |||
Db  66064 GCGCCGGTCGTCACCGTCG 66082
```

```
RESULT  32
AX211739
LOCUS       AX211739               125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN
```

```
  Query Match              2.8%;  Score 38.2;  DB 6;  Length 125401;
  Best Local Similarity   45.6%;  Pred. No. 1.3e+02;
  Matches  118;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    193 ggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatccatcgcggcgaaaag 252
          |||| || | |   ||||   ||   | ||     || | ||  | |   || |    |
Db  66232 GGCGCGGAGATCTTCGCGGACGTGGAATGCCCGCCCGGCACCGCCGACGACGCCCCCGAC 66291

Qy    253 cgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaatcgagttcgaacgc 312
          | || ||| ||  | | | || | | ||  |   |  ||| ||  || | | | ||  ||
Db  66292 CACGGACTGCACCCCGCCCTGCTCGACGCGGCCCGGCACGCCGCCATGGCGGTGGACGGC 66351

Qy    313 tacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngcaatctgatccaa 372
            || |||||  || | |     | |  |     | |   |  |   ||    | | |
Db  66352 ACCGTGCCCGTCGCCTGGCACGGCGTCCGGCTGCACGCCGTCGGCGCCACCGCGCTGCGG 66411

Qy    373 tccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccctgccgtcgatgcc 432
          | |  |  |  |  |   |     |   | | |  ||        | |||  | | |
Db  66412 GTCCGCATCCGCCCCACCACGACCGGCACGCTGACCCTCACCGCGGTCGACGTGCACGGC 66471

Qy    433 gagccgttcgccatcttcg 451
          | |||| ||| || | |||
Db  66472 GCGCCGGTCGTCACCGTCG 66490
```

```
RESULT  33
AC007888
LOCUS       AC007888               164035 bp    DNA     linear   INV 28-FEB-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 57B6-57B13, BAC
            clone BACR10P11, complete sequence.
ACCESSION   AC007888
VERSION     AC007888.7  GI:13162483
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 164035)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
```

```
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 57B6-57B13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164035)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 28, 2001 this sequence version replaced gi:7159348.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .164035
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="57B6-57B13"
                     /clone="BACR10P11 (D651)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    44969 a  36147 c  37118 g  45801 t
ORIGIN


  Query Match              2.8%;  Score 38.2;  DB 3;  Length 164035;
  Best Local Similarity   46.4%;  Pred. No. 1.3e+02;
  Matches  121;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy    1072 aaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccg 1131
           || ||  |  |   ||||| ||  ||   |||   |  |||| ||  |  | | | || |
Db   50134 AACAAACACATACGCAAGTCCATCACCATCGTGTTCTATGGCCACTACAACTTTGAGCAG 50193

Qy    1132 attggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgat 1191
             |   |       |  ||  |      || ||| |  | || | | |||   | |  |
Db   50194 GATCTGAAGGCATCGATCGACAGCATCCTGGACGTGATACCCAACATGCCCATCCTGGTG 50253

Qy    1192 ttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaa 1251
             |    |  |||| | ||||   |||  |  |   ||    ||  |||  | ||  |
Db   50254 CTTCAGGATGGCGGTGCCGAGTACGCCTATCCGCCGGTCAACTACGAACGGAACCTAACC 50313

Qy    1252 gacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgt 1311
             |  ||    | |||  |  | | || | |    | |    |  |  | ||    |||
Db   50314 GCCGGCGAGGAGCGCACCGTGCTATTCCAGAGTCTGTCCTTCGATGTGCGGCGAACGCGG 50373

Qy    1312 aaggtgctggaaaccnttgag 1332
             ||| |||| |  || | | |
Db   50374 CAGGAGCTGAATCCCTTGGCG 50394


RESULT  34
AE003452/c
LOCUS       AE003452               305020 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386038 section 1
            of 15, complete sequence.
ACCESSION   AE003452 AE002575
VERSION     AE003452.2  GI:10726992
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 305020)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 305020)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7291191.
FEATURES             Location/Qualifiers
     source          1. .305020
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA            complement(join(<5457. .6299,6837. .>7250))
                     /gene="CG15226"
```

```
                     /product="CT35163"
                     /db_xref="FLYBASE:FBan0015226"
                     /db_xref="FLYBASE:FBgn0034550"
                     /evidence=not_experimental
     gene            complement(<5457. .>7250)
                     /gene="CG15226"
                     /map="57B3-57B4"
                     /db_xref="FLYBASE:FBan0015226"
                     /db_xref="FLYBASE:FBgn0034550"
                     /evidence=not_experimental
     CDS             complement(join(5457. .6299,6837. .7250))
                     /gene="CG15226"
                     /note="CG15226 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015226"
                     /db_xref="FLYBASE:FBgn0034550"
                     /evidence=not_experimental
                     /protein_id="AAF46624.1"
                     /db_xref="GI:7291192"
                     /translation="MNSGQEQIRTAIPWLYILALSAAQPHIPILRPQDETPANTNSAI
                     MKKPQSDEIVNSDQSNNEVREQFLQGILLGLQQPGSPTISPAVLSDFIAQATTTTRKP
                     AVQKDADGSDNDEDVDYIDFKNGSRIKYDLDDVEQPTIGLVLLLTGQAWGQSFSFPVY
                     TVGSIVSALTGGGGGGGGGGGGGGSRPSSSSNSNAQAEYEAYLNEYYSSQWQQYYEA
                     QAAYNAYLYSLYLASLNNNNNNQVTSTTAAPDTTSSTSTTSTANPTSTATPTTSTAIP
                     TSTGSPTTSTAVATQSTTTASPTTSTAVATSDGSDTTSTAVATQTTTTASPTTSTAVA
                     PADTDTTSTAVQNARIQTTPLPYRYVHSSLVNPYDLHRISSNQIYVRGSAVASADAQS
                     EPVYIPYQNIYRRAQPSPYFGYEGKK"
     mRNA            <10107. .>10826
                     /gene="CG15225"
                     /product="CT35162"
                     /db_xref="FLYBASE:FBan0015225"
                     /db_xref="FLYBASE:FBgn0034551"
                     /evidence=not_experimental
     gene            <10107. .>10826
                     /gene="CG15225"
                     /map="57B3-57B4"
                     /db_xref="FLYBASE:FBan0015225"
                     /db_xref="FLYBASE:FBgn0034551"
                     /evidence=not_experimental
     CDS             10107. .10826
                     /gene="CG15225"
                     /note="CG15225 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015225"
                     /db_xref="FLYBASE:FBgn0034551"
                     /evidence=not_experimental
                     /protein_id="AAF46625.1"
                     /db_xref="GI:7291193"
                     /translation="MPSVNQFKANTLFTMSTKLVLFLCLALFGAMAMAIPLEENLEQN
                     SEVENSQELNSVEENDAEADPSTWRKLFGSEGSAGNLVVSFPQTTTTTTKKSKTTKRP
                     YYPYPPKYPPYPPYPYPHYSPYPAPPYPYPGYYPPPPPPYPYPYPGHGGHSGSGGHDSGH
                     GGHHHHTTTTTTTTTTTKKPNHGQYPPPPPPPPYYPPYPYYPPPPPPPPLPPPSNDSG
                     ESSETSAKICKKFFGIGLICT"
     mRNA            complement(join(38408. .39904,40334. .42109,42444. .42889))
                     /gene="insc"
                     /product="CT9929"
                     /db_xref="FLYBASE:FBan0011312"
                     /db_xref="FLYBASE:FBgn0011674"
     gene            complement(<38408. .>42889)
                     /gene="insc"
                     /note="CG11312"
                     /map="57B4-57B5"
                     /db_xref="FLYBASE:FBan0011312"
                     /db_xref="FLYBASE:FBgn0011674"
     CDS             complement(join(39350. .39904,40334. .42109,42444. .42692))
                     /gene="insc"
                     /note="insc gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0011312"
                     /db_xref="FLYBASE:FBgn0011674"
                     /protein_id="AAF46626.1"
                     /db_xref="GI:7291194"
                     /translation="MSFQRSYSKVWWGSEQSSLTSGGFYSSFSGGSSTSSTSNYPQRQ
                     PASLGPLGNLSCIQEVEDEQNSSKLSWHKHDSPGSLRSQDSGFSDNEESHQLQHQQNL
                     SPPSPQSDKQQTPAATPTSVHSVATPPTVVRRVGNSSFYSPLISVTRRISFTSGQITP
                     KTRSGDEDAEDAELEASRSRLFEQLDSMKLDDGEAEAEEDSPPRPAIRRRRRLMRKPM
                     PEPEAASASEEEEVVIAPPPPPPYNNETVYLGEAPPPYNNETVTFGATEQDESLQQQP
                     SPLRVRSLRFTASTSTPKSGSKIAKRGKKHPEPVASWMSEQRWAGEPEVMCTLQHKSI
                     AQEAYKNYTITTSAVCKLVRQLQQQALSLQVHFERSERVLSGLQASSLPEALAGATQL
                     LSHLDDFTATLERRGVFFNDAKIERRRYEQHLEQIRTVSKDTRYSLERQHYINLESLL
                     DDVQLLKRHTLITLRLIFERLVRVLVISIEQSQCDLLLRANINMVATLMNIDYDGFAS
                     LSDAFVQNEAVRTLLVVVLDHKQSSVRALALRALATLCCAPQAINQLGSCGGIEIVRD
                     ILQVESAGERGAIERREAVSLLAQITAAWHGSEHRVPGLRDCAESLVAGLAALLQPEC
                     CTQTLLLCAAALNNLSRMEATSHYSIMSNEAIFRLIDTLEHQTCGTSVFLYEQIVGML
                     HNMSLNKKCHSHLANGVIINFITSVYQTEFYQTYGSRAESDAQRRTIKTILHTLTRLA
                     SDSPTLGAELLEQWHLPDLLRQSLALKPANSQQQQHLDSSYGGDISQLARQLLRAHGQ
                     EQQLLQATPAPGGSFSSGHQTPEAQTQTNSSGMSSSGSKTKAKSSASPLLKFNLTRQE
                     SFV"
     mRNA            complement(join(44601. .47535,48926. .49299))
                     /gene="sktl"
                     /product="CT28111"
                     /db_xref="FLYBASE:FBan0009985"
                     /db_xref="FLYBASE:FBgn0016984"
     gene            complement(<44601. .>49299)
                     /gene="sktl"
                     /note="CG9985"
                     /map="57B4-57B5"
                     /db_xref="FLYBASE:FBan0009985"
                     /db_xref="FLYBASE:FBgn0016984"
     CDS             complement(45134. .47512)
                     /gene="sktl"
                     /note="sktl gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0009985"
                     /db_xref="FLYBASE:FBgn0016984"
                     /protein_id="AAF46627.2"
                     /db_xref="GI:10726993"
                     /translation="MDTRVELELEPVGKQDRLKDQPDDENEPLENQLELEQEHKSLLT
                     PVKIPPAQTASPDQEATPSTPPPMSLQPGTSRQLFRDAAMAHGGQETTSLLQQELNNI
                     AATQTPASKLRSASSTLDASISRNPSTTGGKHEKKLGHRRVAEGGEVTYKKIQSKQIM
                     GSIQLGIQHTVGSLASKPKRDLLMNDFWEMETISFPPDGSSITPAHHYNDFRFKVYAP
                     IAFRYFRDLFGIAPDDFLMSMCASPLRELSNPGASGSIFYLTTDDEFIIKTVQKKECE
                     FLQKLLPGYYMNLSQNPRTLLPKFFGLYCFHYNSKNVRLVAMNNLLPSDIKMHCKYDL
                     KGSSFRRKASKAERQKASPTFKDLDFAEHHPNGIFLETDKYNALMSTIKRDCMVLESF
                     QIMDYSLLVGIHNLDLAAKEKREERILNARAKLQRKESAAQGPPSLNPDDDAPEADQN
                     QLQAVASYASIPGTSAGAALNRTRSMNRQRLVAHSTALESITADMDVPLEEDEDVPAG
                     GIPARSENDERLILYIGIIDILQSYRLEKKLEHTFKSILYNGDTVSVCRPSFYAKRFQ
                     DAMGKQVFKKTPTFPLKHSPSKRKTSTTQLRSPASRLPPLSTPTNAVRQPIMTMSGMS
                     TPPPAFDDISEEDITAASTSSLQQQRSSNQSNNNRGETEVALREGNNTRGGSGEPSST
                     YHTQYSYDSSGRTGSALTSDYSDDESTGTDLSSRLPKLRTHRLTKTSVQVTTVGYVEG
                     VDTPPKGGQDVVHADVNGKTTMTTTTMLTTTKTAHIQAPSYTSTLVLNDLPR"
     mRNA            complement(join(65140. .66115,66254. .66376,66445. .66631,
                     66701. .>67062))
                     /gene="CG17999"
                     /product="CT40238"
                     /db_xref="FLYBASE:FBan0017999"
                     /db_xref="FLYBASE:FBgn0034552"
     gene            complement(<65140. .>67062)
                     /gene="CG17999"
                     /map="57B5-57B6"
                     /db_xref="FLYBASE:FBan0017999"
                     /db_xref="FLYBASE:FBgn0034552"
     CDS             complement(join(65140. .66115,66254. .66376,66445. .66631,
                     66701. .67046))
```

```
  Query Match              2.8%;  Score 38.2;  DB 3;  Length 305020;
  Best Local Similarity   46.4%;  Pred. No. 1.3e+02;
  Matches  121;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy   1072 aaaaacaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccg 1131
          || ||  |  |  ||||| ||  ||   |||   |  |||| ||  |  | | | || |
Db 203151 AACAAACACATACGCAAGTCCATCACCATCGTGTTCTATGGCCACTACAACTTTGAGCAG 203092

Qy   1132 attggtacttacgagcgcgtaatgccgctagacatcctgcctaccctgcttttgcgcgat 1191
            |  |       |  ||  |    || ||| |  | || | | |||   | |  |
Db 203091 GATCTGAAGGCATCGATCGACAGCATCCTGGACGTGATACCCAACATGCCCATCCTGGTG 203032
```

```
Qy   1192 ttaatcgtcggcgataccgacagcgcgcaagcattgggttgcttggaattggacgaagaa 1251
          |    |  |||| | ||||   |||  |  |    ||    ||  |||  | ||  |
Db 203031 CTTCAGGATGGCGGTGCCGAGTACGCCTATCCGCCGGTCAACTACGAACGGAACCTAACC 202972

Qy   1252 gacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatanggcccgctgttgcgt 1311
          | |  ||    | ||| | |  | | || | |    | |    | |  | ||  |||
Db 202971 GCCGGCGAGGAGCGCACCGTGCTATTCCAGAGTCTGTCCTTCGATGTGCGGCGAACGCGG 202912

Qy   1312 aaggtgctggaaaccnttgag 1332
          ||| ||||  |  || | | |
Db 202911 CAGGAGCTGAATCCCTTGGCG 202891


RESULT  35
AC110159
LOCUS       AC110159               69283 bp    DNA     linear   HTG 10-FEB-2002
DEFINITION  Mus musculus clone RP23-4P22, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC110159
VERSION     AC110159.1  GI:18643511
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 69283)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-4P22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69283)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L17557
                Center clone name: 4_P_22
            --------------.
            * NOTE: This record contains 89 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      660: contig of 660 bp in length
            *      661 760: gap of      100 bp
            *      761     1447: contig of 687 bp in length
            *     1448 1547: gap of      100 bp
            *     1548     2214: contig of 667 bp in length
            *     2215 2314: gap of      100 bp
            *     2315     3006: contig of 692 bp in length
            *     3007 3106: gap of      100 bp
            *     3107     3789: contig of 683 bp in length
            *     3790 3889: gap of      100 bp
            *     3890     4564: contig of 675 bp in length
            *     4565 4664: gap of      100 bp
            *     4665     5337: contig of 673 bp in length
            *     5338 5437: gap of      100 bp
            *     5438     6112: contig of 675 bp in length
            *     6113 6212: gap of      100 bp
            *     6213     6900: contig of 688 bp in length
            *     6901 7000: gap of      100 bp
            *     7001     7664: contig of 664 bp in length
            *     7665 7764: gap of      100 bp
            *     7765     8468: contig of 704 bp in length
            *     8469 8568: gap of      100 bp
            *     8569     9264: contig of 696 bp in length
            *     9265 9364: gap of      100 bp
            *     9365    10059: contig of 695 bp in length
            *    10060 10159: gap of      100 bp
            *    10160    10835: contig of 676 bp in length
            *    10836 10935: gap of      100 bp
            *    10936    11619: contig of 684 bp in length
            *    11620 11719: gap of      100 bp
            *    11720    12389: contig of 670 bp in length
            *    12390 12489: gap of      100 bp
            *    12490    13175: contig of 686 bp in length
            *    13176 13275: gap of      100 bp
            *    13276    13956: contig of 681 bp in length
            *    13957 14056: gap of      100 bp
            *    14057    14736: contig of 680 bp in length
            *    14737 14836: gap of      100 bp
            *    14837    15520: contig of 684 bp in length
            *    15521 15620: gap of      100 bp
            *    15621    16312: contig of 692 bp in length
            *    16313 16412: gap of      100 bp
            *    16413    17082: contig of 670 bp in length
            *    17083 17182: gap of      100 bp
            *    17183    17856: contig of 674 bp in length
            *    17857 17956: gap of      100 bp
            *    17957    18646: contig of 690 bp in length
            *    18647 18746: gap of      100 bp
            *    18747    19392: contig of 646 bp in length
            *    19393 19492: gap of      100 bp
            *    19493    20174: contig of 682 bp in length
            *    20175 20274: gap of      100 bp
            *    20275    20950: contig of 676 bp in length
            *    20951 21050: gap of      100 bp
            *    21051    21716: contig of 666 bp in length
            *    21717 21816: gap of      100 bp
            *    21817    22489: contig of 673 bp in length
            *    22490 22589: gap of      100 bp
            *    22590    23251: contig of 662 bp in length
            *    23252 23351: gap of      100 bp
            *    23352    24010: contig of 659 bp in length
            *    24011 24110: gap of      100 bp
            *    24111    24809: contig of 699 bp in length
            *    24810 24909: gap of      100 bp
            *    24910    25569: contig of 660 bp in length
            *    25570 25669: gap of      100 bp
            *    25670    26354: contig of 685 bp in length
            *    26355 26454: gap of      100 bp
```

```
*    26455    27133: contig of 679 bp in length
*    27134 27233: gap of      100 bp
*    27234    27917: contig of 684 bp in length
*    27918 28017: gap of      100 bp
*    28018    28693: contig of 676 bp in length
*    28694 28793: gap of      100 bp
*    28794    29441: contig of 648 bp in length
*    29442 29541: gap of      100 bp
*    29542    30220: contig of 679 bp in length
*    30221 30320: gap of      100 bp
*    30321    30989: contig of 669 bp in length
*    30990 31089: gap of      100 bp
*    31090    31777: contig of 688 bp in length
*    31778 31877: gap of      100 bp
*    31878    32553: contig of 676 bp in length
*    32554 32653: gap of      100 bp
*    32654    33347: contig of 694 bp in length
*    33348 33447: gap of      100 bp
*    33448    34140: contig of 693 bp in length
*    34141 34240: gap of      100 bp
*    34241    34894: contig of 654 bp in length
*    34895 34994: gap of      100 bp
*    34995    35675: contig of 681 bp in length
*    35676 35775: gap of      100 bp
*    35776    36461: contig of 686 bp in length
*    36462 36561: gap of      100 bp
*    36562    37254: contig of 693 bp in length
*    37255 37354: gap of      100 bp
*    37355    38039: contig of 685 bp in length
*    38040 38139: gap of      100 bp
*    38140    38810: contig of 671 bp in length
*    38811 38910: gap of      100 bp
*    38911    39587: contig of 677 bp in length
*    39588 39687: gap of      100 bp
*    39688    40346: contig of 659 bp in length
*    40347 40446: gap of      100 bp
*    40447    41133: contig of 687 bp in length
*    41134 41233: gap of      100 bp
*    41234    41918: contig of 685 bp in length
*    41919 42018: gap of      100 bp
*    42019    42721: contig of 703 bp in length
*    42722 42821: gap of      100 bp
*    42822    43518: contig of 697 bp in length
*    43519 43618: gap of      100 bp
*    43619    44313: contig of 695 bp in length
*    44314 44413: gap of      100 bp
*    44414    45093: contig of 680 bp in length
*    45094 45193: gap of      100 bp
*    45194    45877: contig of 684 bp in length
*    45878 45977: gap of      100 bp
*    45978    46659: contig of 682 bp in length
*    46660 46759: gap of      100 bp
*    46760    47436: contig of 677 bp in length
*    47437 47536: gap of      100 bp
*    47537    48205: contig of 669 bp in length
*    48206 48305: gap of      100 bp
*    48306    48967: contig of 662 bp in length
*    48968 49067: gap of      100 bp
*    49068    49744: contig of 677 bp in length
*    49745 49844: gap of      100 bp
*    49845    50522: contig of 678 bp in length
*    50523 50622: gap of      100 bp
*    50623    51321: contig of 699 bp in length
*    51322 51421: gap of      100 bp
*    51422    52103: contig of 682 bp in length
*    52104 52203: gap of      100 bp
*    52204    52897: contig of 694 bp in length
*    52898 52997: gap of      100 bp
*    52998    53668: contig of 671 bp in length
*    53669 53768: gap of      100 bp

Query Match             2.8%;  Score 38;  DB 2;  Length 69283;
Best Local Similarity   34.0%;  Pred. No. 1.5e+02;

  Matches  126;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;

Qy     304 ttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnngnngcaat 363
           ||  | | ||| || |   |  |||   || || |||   |     || | | ||
Db   59037 TTAAACCACTAGGCACATAATTCGTGATCAGACATAAATTGGAGCAAAATANNNNNNNNN 59096

Qy     364 ctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatccctgcc 423
                                           |
Db   59097 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 59156

Qy     424 gtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctngcggca 483
                                          | ||| ||  |    ||  | | ||
Db   59157 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGAGATTGATTCTCGCCCTGTGGTGGAT 59216

Qy     484 gaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctggtattg 543
             |   |||  |      || |||  |   | ||   |||         |     ||
Db   59217 TCCTAATTGGAAGCAGGTTAAAAAGTGGTTTAAGAGGCTAGAGAGATGGATCAATGGTTC 59276

Qy     544 agccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgccgtct 603
           ||     || | |  | |  ||   |  | ||  ||       || |  | ||| |  ||
Db   59277 AGATCACTGGCTGCTCTTCCAAAGAACTCGGGTCTGGTCCTCAGCACCCATGTGGCAGCT 59336

Qy     604 gaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtttgagt 663
            | |||  |   ||  || |  | || |   || |  |||| | || ||||  ||  |
Db   59337 CACAATCATCTGAAACTCCAGTCTCAGGGGATCTGATGCCCTCTTCTGGCCTTTTAAACC 59396

Qy     664 ggcacgcacat 674
             | ||||||
Db   59397 ACTAGGCACAT 59407


RESULT  36
AF429315
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
```

```
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT     29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN
```

```
  Query Match             2.8%;  Score 38;  DB 9;  Length 125020;
  Best Local Similarity   10.0%;  Pred. No. 1.5e+02;
  Matches   98;  Conservative  413;  Mismatches  462;  Indels    9;  Gaps    3;

Qy     180 caaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgccatcca 239
           ||:: :   :  :::: : ::::  :::::::: : : :: ::  :::::|:  :  ::: 
Db   16872 CARRKMMKSHAGRRRMCYYSKSWMSMBMSVSYSVKMHSMHASBSCMHWBKCMTWSCCMSM 16931

Qy     240 tcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacgaaat 299
           ::::  |  :  |  | :  : :::::: :: :   :: |  ::::::::|  :  |
Db   16932 MYKSSWWGSSWGMCCWGRRRSKGWKWYSRGMSRSKSMRYTGGSKMMRSSMMCTSSCYASM 16991

Qy     300 cgagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnng-nn 358
           | :  ::   :::  | |:::|  :::    :| : ::: :: : ::: : :  :
Db   16992 CCMCCWSCCCMRSCCCMCRSYCCCMRYCCACKCYMSSYWTMSASYMSRRSYWKRSMKCWS 17051

Qy     359 gcaatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaatcc 418
           ::   : : :|  : | |:  :|:|| ::| : :: :   : |     ::::::: :::: 
Db   17052 RMCSRSRSSKCKSRGCGGSMGKGKGGKSYGRKTKRKSRGKMKGAKWYMYYRRSMRKMKMY 17111

Qy     419 ctgccgtcgatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctng 478
           : : :  :  :  ::||:  : :  :  :  :|  ::  ::  : :::  ::
Db   17112 SSKGMYCMYCWCWGRRGCYCSCMTSRSAMCCSYCYAKCKSMCYSCYYGMSMKGYYYACSY 17171

Qy     479 cggcagaccctgtggttgtgatcaaagaagccgncgangatttcagacgangtntgctgg 538
           |:: : :|: |    |: |  |       || |    |             : | :: :
Db   17172 RGSMSSKYCMRGSTYSTSCGCCCTTTTTCCCCCNANTGGGGAAGCTTTTNCNKKTYSYYR 17231

Qy     539 tattgagccgtttgaccgagcgtaaaatccatgtgtgtaaggcagctggcgcagacgtgc 598
           ::   | |    :   : :::  | :     ::  |  :|: |: :   : ::: :  :
Db   17232 KRNGCAMCKYNNYYNNSWRSSCRAGMSCTKYKSSMTMSMASYCWCMSMYYCSMRSMASRG 17291

Qy     599 cgtctgaaaatgctgccaacatcgaaacacatgaattcggcggcccgcatccggccggtt 658
           ::::  : :::::::::  :  |  : : :   :   ::|:: ::: :::::::: |  |
Db   17292 MSWSSYMKMMKSSWRMSCYMKCCWCMKMCYCMMRSMRSRGMSYYMYASWKSSSRGCYCTR 17351

Qy     659 tgagtggcacgcacattcatttcattgagccggtcggtgcaaacaaaaccgtttggacca 718
             :     ::: : ::: :  :: :::: ::      :::    :  |::: ::: :|::
Db   17352 CYW-----CMSSKSCYKSYYMMMRSKRMKGMKWSRCWGSMWGSASRSSCYKCYKSMRCSM 17406

Qy     719 tcaattatcaagatgtaattgccatcggacgtttgtttgcaacaggccgtctgaacaccg 778
           :   ::      : :      :  :: ::  : ::  ::::::::|    :  : | :
Db   17407 MSSKCYRCAGCMMMKGGYMRYMRCWSMKRRWWGKSAMYRMRRMWKGRGAMMMCMKCYSRM 17466

Qy     779 agcgcgtgattgctttgggtggttctcaagtcaacaaaccacgcctcttgcgtaccgttt 838
           :| :  : :  :       :||:::||   |: : :  |:| :: ::::|  :| |:
Db   17467 SRCMMMMKSYCASCRSCAMMSGTYMKCASYMCMSSYWCTCWCTSYMSYYRCTCWCKGWSS 17526

Qy     839 tgggtgcgaaagtatcgcaaattactgcgggcgaattggttgacgcagacaaccgcgtga 898
           : :::: :  ::  : | :  : :|  ::| :: :    : :  ::::|  :: : :|:
Db   17527 YTKSKSWSSSMSSYKKGRRKSYSMCCTSRGAMSCWRRCCYMRGASSMRAGSMSRRRAKGR 17586

Qy     899 tttccggttcggtat---tgaacggcgcgattacacaaggcgcgcacgattatttgggac 955
           :  :::::| ::   :   | | :: :|::| |::::: : :::       :::
Db   17587 SWGGRSKWWMTGGMWRSKYYYYCTGRRMMMTYMCCCWRRRSYMYYRSMSAMGMRKSSWSG 17646

Qy     956 gctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctggg 1015
             : :|: : | :|   :    :  |::|   ::    |:: ::|::: : ::|: :::
```

```
Db   17647 MRMMGSASSRRCKSASRSSWCSRRMKGMRGSCWSSKMWGGSRSRSASSKCKGSRGMRRRS 17706

Qy    1016 ttgcgccgcagccggacaaatactccatcacgcgtacgaccctcggccatttcctgaaaa 1075
           : :::: :  :  :: :   : ::   :|: ::| : : |:::: ::|::  :::: :
Db   17707 KRSSKYRKRGRGKKRSMTKSKGSKGKSKCWKRSGSMTSSCYYYSASSCMWMMSSKSCMCC 17766

Qy    1076 acaaactcttcaagttcacgacagccgtcaacggtggcgaccgcgccatggtgccgattg 1135
             :::   |  |  :  | :       |    :     |: | |:: ::|  : ::::| :
Db   17767 CMMMKRRCACCYSSMSCTSMYRCCWGGMKSYSTCGTTCKKCTGYKKSRTWYWWMYSWTSK 17826

Qy    1136 gtacttacgagcgcgtaatgcc 1157
           ::   || : | ::|::  |:
Db   17827 MYMWRAACMMYCYMSTRRMMCM 17848
```

```
RESULT  37
AF435865
LOCUS       AF435865               9743 bp    RNA     linear   VRL 14-NOV-2001
DEFINITION  Rubella virus isolate Ulrike, complete genome.
ACCESSION   AF435865
VERSION     AF435865.1  GI:16923733
KEYWORDS    .
SOURCE      Rubella virus.
  ORGANISM  Rubella virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Rubivirus.
REFERENCE   1  (bases 1 to 9743)
  AUTHORS   Hofmann,J., Renz,M., Meyer,S., von Haeseler,A. and Liebert,U.G.
  TITLE     Rubella virus (isolate 'Ulrike') genomic RNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9743)
  AUTHORS   Hofmann,J., Renz,M., Meyer,S., von Haeseler,A. and Liebert,U.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2001) Institute of Virology, University of
            Leipzig, Max-Buerger-Forschungszentrum, Johannisallee 30, Leipzig
            04103, Germany
FEATURES             Location/Qualifiers
     source          1. .9743
                     /organism="Rubella virus"
                     /isolate="Ulrike"
                     /db_xref="taxon:11041"
     CDS             22. .6372
                     /codon_start=1
                     /product="nonstructural proteins NSP1 and NSP2"
                     /protein_id="AAL31565.1"
                     /db_xref="GI:16923734"
                     /translation="MEKLLDEVLAPGGPYNLTVGSWVRDHVRSIVEGAWEVRDVVSAA
                     QKRAIVAVIPRPVFTQMQVSDHPALHAISRYTRRHWIEWGPKEALHVLIDPSPGLLRE
                     VARVERRWVALCLHRTARKLATALAETASEAWHADYVCALRGAPSGPFYVHPEDVPHG
                     GRAVADRCLLYYTPMQMCELMRTIDATLLVAVDLWPVALAAHVGDDWDDLGIAWHLDH
                     DGGCPADCRGAGAGPTPGYTRPCTTRIYQVLPDTAHPGRLYRCGPRLWTRDCAVAELS
                     WEVAQHCGHQARVRAVRCTLPIRHVRSLQPSARVRLPDLVHLAEVGRWRWFSLPRPVF
                     QRMLSYCKTLSPDAYYSERVFKFKNALSHSITLAGNVLQEGWKGTCAEEDALCAYVAF
                     RAWQSNARLAGIMKSAKRCAADSLSVAGWLDTIWGAIKRFFGSVPLAERMEEWEQDAA
                     VAAFDRGPLEDGGRHLDTVQPPKSPPRPEIAATWIVHAASADRHCACAPRCDVPRERP
                     SAPACPPDDEALIPPWLFAERRALRCREWDFEALRARADTAAAPAPLAPRPARYPTVL
                     YRHPAHHGPWLTLDEPGEADAALVLCDPLGQPLRGPERHFAAGAHMCAQARGLQAFVR
                     VVPPPERPWADGGARAWAKFFRGCAWAQRLLGEPAVMHLPYTDGDVPQLIALALRTLA
                     QQGAALALSVRDLPRGAAFDANAVTAAVRAGPGQFAATSPPPGDPPPPRRARRSQRHS
                     DARGTPPPAPARDPPPPAPSPPAPPRAGDPDSPTSAEPADRARHAELEVVYEPSGPPT
                     STKADPDSDIVESYARAAGPVHLRVRNIMDPPPGCKVVVNAANEGLLAGSGVCGAIFA
                     NATAALAADCRRLAPCPTGEAVATPGHGCGYTHIIHAVAPRRPRDHAALEEGEALLER
                     AYRSIVALAAARRWACVACPLLGAGVYGWSAAESLRAALAATRTEPAERVSLHICHPD
                     RATLTHASVLVGAGLAARRVSPPPTEPLASCPAGDLGRPAQRSASPPATPLGDATAPE
                     PRGCQGCELCRCTRVTNDRAYVNLWLERDRGATSWAMRIPEVVVYGPEHLATHFPLNH
                     YSVLKPAEVRPPRGMCGSDMWRCRGWQGMPQVRCTPSNAHAALCRTGVPPRVSTRGGE
                     LDPNTCWLRAAANVAQAARACGAYTSAGCPKCAYGRALSEARTHEDFAALSQRWSASH
                     ADASPDGTGDPLDPLMETVGCACSRVWVGTEHEAPPDHLLVSLHRAPNGPWGVVLEVR
                     ARPEGGNPTGHFVCAVGGGPRRVSDRPHLWLAVPLSRGGGTCAATDEGLAQAYYDDLE
                     VRRLGDDAMARAALASVQRPRKGPYNIRVWNMAAGAGKTTRILAAFTREDLYVCPTNA
                     LLHEIQAKLRARDIDIKNAATYERALTKPLAAYRRIYIDEAFTLGGEYCAFVASQTTA
                     EVICVGDRDQCGPHYANNCRTPVPDRWPTERSRHTWRFPDCWAARLRAGLDYDIEGER
                     TGTFACNLWDGRQVDLHLAFSRETVRRLHEAGIRAYTVREAQGMSVGTACIHVGRDGT
```

```
                     DVALALTRDLAIVSLTRASDALYLHELEDGSLRAAGLSAFLDAGALAELKEVPAGIDR
                     VVAVEQAPPPLPPADGIPEAQDVPPFCPRTLEELVFGRAGHPHYADLNRVTEGEREVR
                     YMRISRHLLNKNHTEMPGTERVLSAVCAVRRYRAGEDGSTLRTAVARQHPRPFRQIPP
                     PRVTAGVAQEWRMTYLRERIDLTDVYTQMGVAARELTDRYARRYPEIFAGMCTAQSLS
                     VPAFLKATLKCVDAALGPRDTEDCHAAQGKAGLEIRAWAKEWVQVMSPHFRAIQKIIM
                     RALRPQFLVAAGHTEPEVDAWWQAHYTTNAIEVDFTEFDMNQTLATRDVELEISAALL
                     GLPCAEDYRALRAGSYCTLRELGSTETGCERTSGEPATLLHNTTVAMCMAMRMVPKGV
                     RWAGIFQGDDMVIFLPEGARSAALKWTPAEVGLFGFHIPVKHVSTPTPSFCGHVGTAA
                     GLFHDVMHQAIKVLCRRFDPDVLEEQQVALLDRLRGVYAALPDTVAANAAYYDYSAER
                     VLAIVRELTAYARGRGLDHPATIGALEEIQTPYARANLHDAD"
     CDS             6493. .9684
                     /codon_start=1
                     /product="structural proteins C, E2 and E1"
                     /protein_id="AAL31566.1"
                     /db_xref="GI:16923735"
                     /translation="MASTTPITMEDLQKALEAQSRALRAELAAGASQSRRPRPPRQRD
                     SSTSGDDSGRDSGGPRRRRGNRGRGQRKDWSMAPPPPEERQESRSQTPAPKPSRAPPQ
                     QPQPPRMQTGRGGSAPRPELGPPTNPFQAAVARGLRPPLHDPDTEAPTEACVTSWLWS
                     EGEGAVFYRVDLHFTNLGTPPLDEDGRWDPALMYNPCGPEPPAHVVRAYNQPAGDVRG
                     VWGKGERTYTEQDFRVGGTRWHRLLRMPVRGLDGDSAPLPPHTTERIETRSARHPWRI
                     RFGAPQAFLAGLLLAAVAVGTARAGLQPRADMAAPPAPPQPPCAHGQHYGHHHHQLPF
                     LGHDGHHGGTLRVGQHHRNASDVLPGHWLQGGWGCYNLSDWHQGTHVCHTKHMDFWCV
                     EHDRPPPATPTPLTTAANSTTAATPATAPAPCHAGLNDSCGGFLSGCGPMRLRHGADT
                     RCGRLICGLSTTAQYPPTRFACAMRWGLPPWELVVLTARPEDGWTCRGVPAHPGTRCP
                     ELVSPMGRATCSPASALWLATANALSLDHALAAFVLLFPWVLIFMVCRRACRRRGAAA
                     ALTAVVLQGYNPPAYGEEAFTYLCTAPGCATQAPVPVRLAGVRFESKIVDGGCFAPWD
                     LEATGACICEIPTDVSCEGLGAWVPTAPCARIWNGTQRACTFWAVNAYSSGGYAQLAS
                     YFNPGGSYYKQYHPTACEVEPAFGHSDAACWGFPTDTVMSVFALASYVQHPHKTVRVK
                     FHTETRTVWQLSVAGVSCNVTTEHPFCNTPHGQLEVQVPPDPGDLVEYIMNYTDNQQS
                     RWGLGSPNCHGPDWASPVCQRHSPDCSRLVGATPERPRLRLVDADDPLLRTAPGPGEV
                     WVTPVIGSQARKCGLHIRAGPYGHATVEMPEWIHAHTTSDPWHPPGPLGLKFKTVRPV
                     ALPRALAPPRNVRVTGCYQCGTPALVEGLAPGGGNCHLTVNGEDVGAVPPGKFVTAAL
                     LNTPPPYQVSCGGESDRASARVIDPAAQSFTGVVYGTHTTAVSETRQTWAEWAAAHWW
                     QLTLGAVCALLLAGLLACCAKCLYYLRGAIAPR"
BASE COUNT     1448 a   3761 c   3007 g   1527 t
ORIGIN


  Query Match             2.8%;  Score 37.2;  DB 14;  Length 9743;
  Best Local Similarity   45.7%;  Pred. No. 2.2e+02;
  Matches  123;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy      55 gtcatttatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggt 114
           | ||  | || || | ||| |        || |  |  |     ||| ||    |  |
Db    3925 GGCACCTGTGCCGCGACCGACGAGGGGCTAGCCCAGGCGTACTACGACGACCTCGAGGTG 3984

Qy     115 atgcgccccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgttt 174
             |||| |  |    | | || ||   | |  ||| ||  |  || ||||   |
Db    3985 CGCCGCCTCGGGGATGACGCCATGGCCCGGGCGGCCCTCGCATCAGTCCAACGCCCTCGC 4044

Qy     175 gaagacaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgcc 234
           ||| |    |  ||    | ||   |  |  || | |   | ||||| | | ||  |
Db    4045 AAAGGCCCCTACAATATCAGGGTATGGAACATGGCCGCAGGCGCCGGCAAGACCACCCGC 4104

Qy     235 atccatcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgac 294
           ||||     | |   | ||||| |   |  |  |||      ||  ||    |   ||
Db    4105 ATCCTCGCTGCCTTCACGCGCGAAGACCTCTACGTCTGCCCCACCAATGCGCTCCTGCAC 4164

Qy     295 gaaatcgagttcgaacgctacgcgcccga 323
           || ||| ||  | |||    ||||| |||
Db    4165 GAGATCCAGGCCAAACTTCGCGCGCGCGA 4193


RESULT  38
AF435866
LOCUS       AF435866               9743 bp    RNA     linear   VRL 14-NOV-2001
DEFINITION  Rubella virus isolate Surkova, complete genome.
ACCESSION   AF435866
VERSION     AF435866.1  GI:16923736
KEYWORDS    .
SOURCE      Rubella virus.
  ORGANISM  Rubella virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Rubivirus.
REFERENCE   1  (bases 1 to 9743)
  AUTHORS   Hofmann,J., Renz,M., Meyer,S., von Haeseler,A. and Liebert,U.G.
  TITLE     Rubella virus (isolate 'Surkova') genomic RNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9743)
  AUTHORS   Hofmann,J., Renz,M., Meyer,S., von Haeseler,A. and Liebert,U.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2001) Institute of Virology, University of
            Leipzig, Max-Buerger-Forschungszentrum, Johannisallee 30, Leipzig
            04103, Germany
FEATURES             Location/Qualifiers
     source          1. .9743
                     /organism="Rubella virus"
                     /isolate="Surkova"
                     /db_xref="taxon:11041"
     CDS             22. .6372
                     /codon_start=1
                     /product="nonstructural proteins NSP1 and NSP2"
                     /protein_id="AAL31567.1"
                     /db_xref="GI:16923737"
                     /translation="MEKLLDEVLAPGGPYNLTVGSWVRDHVRSIVEGAWEVRDVVSAA
                     QKRAIVAVIPRPVFTQMQVSDHPALHAISRYTRRHWIEWGPKEALHVLIDPSPGLLRE
                     VARVERRWVALCLHRTARKLATALAETASEAWHADYVCALRGAPSGPFYVHPEDVPHG
                     GRAVADRCLLYYTPMQMCELMRTIDATLLVAVDLWPVALAAHVGDDWDDLGIAWHLDH
                     DGGCPADCRGAGAGPTPGYTRPCTTRIYQVLPDTAHPGRLYRCGPRLWTRDCAVAELS
                     WEVAQHCGHQARVRAVRCTLPIRHVRSLQPSARVRLPDLVHLAEVGRWRWFSLPRPVF
                     QRMLSYCKTLSPDAYYSERVFKFKNALSHSITLAGNVLQEGWKGTCAEEDALCAYVAF
                     RAWQSNARLAGIMKSAKRCAADSLSVAGWLDTIWGAIKRFFGSVPLAERMEEWEQDAA
                     VAAFDRGPLEDGGRHLDTVQPPKSPPRPEIAATWIVHAASADRHCACAPRCDVPRERP
                     SAPACPPDDEALIPPWLFAERRALRCREWDFEALRARADTAAAPAPLAPRPARYPTVL
                     YRHPAHHGPWLTLDEPGEADAALVLCDPLGQPLRGPERHFAAGAHMCAQARGLQAFVR
                     VVPPPERPWADGGARAWAKFFRGCAWAQRLLGEPAVMHLPYTDGDVPQLIALALRTLA
                     QQGAALALSVRDLPGGAAFDANAVTAAVRAGPGQLAATSPTPGDPPPPRRARRSQRPS
                     DARGTPPPAPVRDPPPPAPSPPAPPRVGDPVPPTTAEPADRARHAELEVVYEPSGPPT
                     PSKADPDSDIVESYARAAGPVHLRVRNIMDPPPGCKVVVNAANEGLLAGSGVCGAIFA
                     NATAALAADCRRLAPCPTGEAVATPGHGCGYTHIIHAVAPRRPRDPAALEEGEALLER
                     AYRSIVALAAARRWARVACPLLGAGVHGWSAAESLRAALAATRAEPAERVSLHICHPD
                     RATLTHASVLVGAGLAARRVSPPPTEPLASCPAGDPGRPAQRSASPPATPLGDATAPE
                     PRGCQGCELCRYTRVTNDRAYVNLWLERDRGATSWAMRIPEVVVYGPEHLATHFPLNH
                     YSVLKPAEVRPPRGMCGSDMWRCRGWQGMPQVRCTPSNAHAALCRTGVPPRVSTRGGE
                     LDPNTCWLRAAANVAQAARACGAYTSAGCPKCAYGRALSEARTHEDFAALSQRWSASH
                     ADASPDGTGDPLDPLMETVGCACSRVWVGTEHEAPPDHLLVSLHRAPNGPWGVVLEVR
                     ARPEGGNPTGHFVCAVGGGPRRVSDRPHLWLAVPLSRGGGTCAATDEGLAQAYYDDLE
                     VRRLGDDAMARAALASVQRPRKGPYNIRVWNMAAGAGKTTRILAAFTREDLYVCPTNA
                     LLHEIQAKLRARDIDIKNAATYERALTKPLAAYRRIYIDEAFTLGGEYCAFVASQTTA
                     EVICVGDRDQCGPHYANNCRTPVPDRWPTERSRHTWRFPDCWAARLRAGLDYDIEGER
                     TGTFACNLWDGRQVDLHLAFSRETVRRLHEAGIRAYTVREAQGMSVGTACIHVGRDGT
                     DVALALTRDLAIVSLTRASDALYLHELEDGSLRAAGLSAFLDAGALAELKEVPAGIDR
                     VVAVEQAPPPLPPADGIPEAQDVPPFCPRTLEELVFGRAGHPHYADLNRVTEGEREVR
                     YMRISRHLLNKNHTEMPGTERVLSAVCAVRRYRAGEDGSTLRTAVARQHPRPFRQIPP
                     PRVTAGVAQEWRMTYLRERIDLTDVYTQMGVAARELTDRYARRYPEIFAGMCTAQSLS
                     VPAFLKATLKCVDAALGPRDTEDCHAAQGKAGLEIRAWAKEWVQVMSPHFRAIQKIIM
                     RALRPQFLVAAGHTEPEVDAWWQAHYTTNAIEVDFTEFDMNQTLATRDVELEISAALL
                     GLPCAEDYRALRAGSYCTLRELGSTETGCERTSGEPATLLHNTTVAMCMAMRMVPKGV
                     RWAGIFQGDDMVIFLPEGARSAALKWTPAEVGLFGFHIPVKHVSTPTPSFCGHVGTAA
                     GLFHDVMHQAIKVLCRRFDPDVLEEQQVALLDRLRGVYAALPDTVAANAAYYDYSAER
                     VLAIVRELTAYARGRGLDHPATIGALEEIQTPYARANLHDAD"
     CDS             6493. .9684
                     /codon_start=1
                     /product="structural proteins C, E2 and E1"
                     /protein_id="AAL31568.1"
                     /db_xref="GI:16923738"
                     /translation="MASTTPITMEDLQKALEAQSRALRAELAAGASQSRRPRPPRQRD
                     SSTSGDDSGRDSGGPRRRRGNRGRGQRKDWSMAPPPPEERQESRSQTPAPKPSRAPPQ
                     QPQPPRMQTGRGGSAPRPELGPPTNPFQAAVARGLRPPLHDPDTEAPTEACVTSWLWS
                     EGEGAVFYRVDLHFTNLGTPPLDEDGRWDPALMYNPCGPEPPAHVVRAYNQPAGDVRG
                     VWGKGERTYTEQDFRVGGTRWHRLLRMPVRGLDGDSAPLPPHTTERIETRSARHPWRI
                     RFGAPQAFLAGLLLAAVAVGTARAGLQPRADMAAPPAPPQPPCAHGQHYGHHHHQLPF
                     LGHDGHHGGTLRVGQHHRNASDVLPGHCLQGGWGCYNLSDWHQGTHVCHTKHMDFWCV
                     EHDRPPPATPTPLTTAANSTTAATPATAPAPCHAGLNDSCGGFLSGCGPMRLRHGADT
                     RCGRLICGLSTTAQYPPTRFACAMRWGLPPWELVVLTARPEDGWTCRGVPAHPGTRCP
                     ELVSPMGRATCSPASALWLATANALSLDHALAAFVLLFPWVLIFMVCRRACRRRGAAA
```

```
                     ALTAVVLQGYNPPAYGEEAFTYLCTAPGCATQTPVPVRLAGVRFESKIVDGGCFAPWD
                     LEATGACICEIPTDVSCEGLGAWVPTAPCARIWNGTQRACTFWAVNAYSSGGYAQLAS
                     YFNPGGSYYKQYHPTACEVEPAFGHSDAACWGFPTDTVMSVFALASYVQHPHKTVRVK
                     FHTETRTVWQLSVAGVSCNVTTEHPFCNTPHGQLEVQVPPDPGDLVEYIMNYTGNQQS
                     RWGLGSPNCHGPDWASPVCQRHSPDCSRLVGATPERPRLRLVDADDPLLRTAPGPGEV
                     WVTPVIGSQARKCGLHIRAGPYGHATVEMPEWIHAHTTSDPWHPPGPLGLKFKTVRPV
                     ALPRALAPPRNVRVTGCYQCGTPALVEGLAPGGGNCHLTVNGEDVGAVPPGKFVTAAL
                     LNTPPPYQVSCGGESDRASARVIDPAAQSFTGVVYGTHTTAVSETRQTWAEWAAAHWW
                     QLTLGAVCALLLAGLLACCAKCLYYLRGAIAPR"
BASE COUNT     1450 a   3760 c   3013 g   1520 t
ORIGIN


  Query Match             2.8%;  Score 37.2;  DB 14;  Length 9743;
  Best Local Similarity   45.7%;  Pred. No. 2.2e+02;
  Matches  123;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy      55 gtcatttatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggt 114
           | || || || ||   ||| |         | |  |  |     ||| ||  |  |
Db    3925 GGCACTTGTGCCGCCACCGACGAGGGGCTGGCCCAGGCGTACTACGACGACCTCGAGGTG 3984

Qy     115 atgcgcccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgttt 174
             |||| |  |    | | || ||   | |  ||| ||  |  || ||||   |
Db    3985 CGCCGCCTCGGGGACGACGCCATGGCCCGGGCGGCCCTCGCATCAGTCCAACGCCCCCGC 4044

Qy     175 gaagacaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgcc 234
           ||| |    |  ||    | ||   |   |  ||  | |   | ||||| | | ||  |
Db    4045 AAAGGCCCCTACAATATCAGGGTATGGAACATGGCCGCAGGCGCTGGCAAGACCACCCGC 4104

Qy     235 atccatcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgac 294
           ||||     | |   | |||||  |   |  |  |||        ||  ||    |    ||
Db    4105 ATCCTCGCTGCCTTCACGCGCGAAGACCTCTACGTCTGCCCCACCAATGCGCTCCTGCAC 4164

Qy     295 gaaatcgagttcgaacgctacgcgcccga 323
           || ||| ||  | ||| |  ||||| |||
Db    4165 GAGATCCAGGCCAAACTCCGCGCGCGCGA 4193


RESULT  39
AX009468
LOCUS       AX009468                9762 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 1 from Patent WO9961637.
ACCESSION   AX009468
VERSION     AX009468.1  GI:9996753
KEYWORDS    .
SOURCE      Rubella virus.
  ORGANISM  Rubella virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Rubivirus.
REFERENCE   1  (bases 1 to 9762)
  AUTHORS   Chantler,J. and Lund,K.
  TITLE     Gene sequences of rubella virus associated with attenuation
  JOURNAL   Patent: WO 9961637-A 1 02-DEC-1999;
            UNIV BRITISH COLUMBIA (CA); CHANTLER JANET (CA); LUND KAREN (CA)
FEATURES             Location/Qualifiers
     source          1. .9762
                     /organism="Rubella virus"
                     /strain="Cendehill"
                     /db_xref="taxon:11041"
BASE COUNT     1467 a   3764 c   3006 g   1525 t
ORIGIN


  Query Match             2.8%;  Score 37.2;  DB 6;  Length 9762;
  Best Local Similarity   45.7%;  Pred. No. 2.2e+02;
  Matches  123;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy      55 gtcatttatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggt 114
           | ||  | || || | ||| |         | |  |  |     ||| ||    |  |
Db    3944 GGCACCTGTGCCGCGACCGACGAGGGGCTGGCCCAGGCGTACTACGACGACCTCGAGGTG 4003

Qy     115 atgcgcccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgttt 174
             |||| |  |    | | || ||   | |  ||| ||  |  || ||||   |
Db    4004 CGCCGCCTCGGGGATGACGCCATGGCCCGGGCGGCCCTCGCATCAGTCCAACGCCCTCGC 4063

Qy     175 gaagacaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgcc 234
           ||| |    |  ||    | ||   |   |  ||  | |   | ||||| | | ||  |
Db    4064 AAAGGCCCTTACAATATCAGGGTATGGAACATGGCCGCAGGCGCTGGCAAGACCACCCGC 4123

Qy     235 atccatcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgac 294
           ||||     | |   | |||||  |   |  |  |||        ||  ||    |    ||
Db    4124 ATCCTCGCTGCCTTCACGCGCGAAGACCTCTACGTCTGCCCCACCAATGCGCTCCTGCAC 4183

Qy     295 gaaatcgagttcgaacgctacgcgcccga 323
           || ||| ||  | ||| |  ||||| |||
Db    4184 GAGATCCAGGCCAAACTCCGCGCGCGCGA 4212


RESULT  40
AF188704
LOCUS       AF188704                9762 bp    RNA     linear   VRL 20-JAN-2000
DEFINITION  Rubella virus complete genome.
ACCESSION   AF188704
VERSION     AF188704.1  GI:6716708
KEYWORDS    .
SOURCE      Rubella virus.
  ORGANISM  Rubella virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Rubivirus.
REFERENCE   1  (bases 1 to 9762)
  AUTHORS   Lund,K.D. and Chantler,J.K.
  TITLE     Mapping of Genetic Determinants of Rubella Virus Associated with
            Growth in Joint Tissue
  JOURNAL   J. Virol. 74 (2), 796-804 (2000)
  MEDLINE   20091334
   PUBMED   10623741
REFERENCE   2  (bases 1 to 9762)
  AUTHORS   Chantler,J.K. and Lund,K.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) Pathology and Lab Medicine, University of
            British Columbia, 950 West 28th Avenue, Vancouver, BC V5Z 4H4,
            Canada
FEATURES             Location/Qualifiers
     source          1. .9762
                     /organism="Rubella virus"
                     /virion
                     /strain="Cendehill"
                     /db_xref="taxon:11041"
     CDS             <1. .6391
                     /codon_start=2
                     /product="nonstructural capsid protein"
                     /protein_id="AAF26709.1"
                     /db_xref="GI:6716709"
                     /translation="NGSYRTSLRTPIPMEKLLDEVLAPGGPYNLTVGSWVRDHVRSIV
                     EGAWEVRDVVTAAQKRAIVAVIPRPVFTQMQVSDHPALHAISRYTRRHWIEWGPKEAL
                     HVLIDPSPGLLREVARVERRWVALCLHRTARKLATALAETASEAWHADYVCALRGAPS
                     GPFYVHPEDVPHGGRAVADRCLLYYTPMQMCELMRTIDATLLVAVDLWPVALAAHVGD
                     DWDDLGIAWHLDHDGGCPADCRGAGAGPTPGYTRPCTTRIYQVLPDTAHPGRLYRCGP
                     RLWTRDCAVAELSWEVAQHCGHQARVRAVRCTLPIRHVRSLQPSARVRLPDLVHLAEV
                     GRWRWFSLPRPVFQRMLSYCKTLSPDAYYSERVFKFKNALSHSITLAGNVLQEGWKGT
                     CAEEDALCAYVAFRAWQSNARLAGIMKGAKRCAADSLSVAGWLDTIWDAIKRFFGSVP
                     LAERMEEWEQDAAVAAFDRGPLEDGGRHLDTVQPPKSPPRPEIAATWIVHAASDRHC
                     ACAPRCDVPRERPSAPAGQPDDEALIPPWLFAERRALRCREWDFEALRARADTAAAPA
                     PLAPRPARYPTVLYRHPAHHGPWLTLDEPGEADAALVLCDPLGQPLRGPERHFAAGAH
                     MCAQARGLQAFVRVVPPPERPWADGGARAWAKFFRGCAWAQRLLGEPAVMHLPYTDGD
                     VPQLIALALRTLAQQGAALALSVRDLPGGAAFDANAVTAAVRAGPGQLAATSPPPGDP
                     PPPRRARRSQRHSDARGTPPPAPVRDPPPPAPSPPAPPRAGDPVPPTPAEPADRARDA
                     ELEVAYEPSGPPTSTKADPDSDIVESYARAAGPVHLRVRDIMDPPPGCKVVVNAANEG
                     LLAGSGVCGAIFANATAALAADCRRLAPCPTGEAVATPGHGCGYTHIIHAVAPRRPRD
                     PAALEEGEALLERAYRSIVALAAARRWAYVACPLLGAGVYGWSAAESLRAALAATRAE
                     PVERVSLHICHPDRATLTHASVLVGAGLAARRVSPPPTEPLASCPAGGPGRPAQRSAS
                     PPATPLGDATAPEPRGCQGCELCRHTRVTNDRAYVNLWLERDRGATSWAMRIPEVVVY
                     GPEHLATHFPLNHYSVLKPAEVRPPRGMCGSDMWRCRGWQGMPQVRCTPSNAHAALCR
                     TGVPPRVSTRGGELDPNTCWLRAAANVAQVARACGAYTSAGCPKCAYGRALSEARTHE
                     DFAALSQRWSASHADASPDGTGDPLDPLMETVGCACSRVWVGSEQEAPPDHLLVSLHR
                     APNGPWGVVLEVRARPEGGNPTGHFVCAVGGGPRRVSDRPHLWLAVPLSRGGGTCAAT
```

```
                     DEGLAQAYYDDLEVRRLGDDAMARAALASVQRPRKGPYNIRVWNMAAGAGKTTRILAA
                     FTREDLYVCPTNALLHEIQAKLRARDIDIKNAATYERALTKPLAAYRRIYIDEAFTLG
                     GEYCAFVASQTTAEVICVGDRDQCGPHYANNCRTPVPDRWPTERSRHTWRFPDCWAAR
                     LRAGLDYDIEGERTGIFACNLWDGRQVDLHLAFSRETVRRLHEAGIRAYTVREAQGMS
                     VGTACIHVGRDGTDVALALTRDLAIVSLTRASDALYLHELEDGSLRAAGLSAFLDAGA
                     LAELKEVPAGIDRVVAVEQAPPPLPPADGIPEAQDVPPFCPRTLEELVFGRAGHPHYA
                     DLNRVTEGEREVRYMRISRHLLNKNHTEMPGTERVLSAVCAVRRYRAGEDGSTLRTAV
                     ARQHPRPFRQIPPPRVTAGVAQEWRMTYLRERIDLTDVYTQMGVAARELTDRYARRYP
                     EIFAGMCTAQSLSVPAFLKATLKCVDAALGPRDTEDCHAAQGKAGLEIRAWAKEWVQV
                     MSPHFRAIQKIIMRALRPQFLVAAGHTEPEVDAWWQAHYTTNAIEVDFTEFDMNQTLA
                     TRDVELEISAALLGLPCAEDYRALRAGSYCTLRELGSTETGCERTSGEPATLLHNTTV
                     AMCMAMRMVPKGVRWAGIFQGDDMVIFLPEGARSAALKWTPAEVGLFGFHIPVKHVST
                     PTPSFCGHVGTAAGLFHDVMHQAIKVLCRRFDPDVLEEQQVALLDRLRGVYAALPDTV
                     AANAAYYDYSAERVLAIVRELTAYARGRGLDHPATIGALEEIQTPYARANLHDAD"
     CDS             6512. .9703
                     /codon_start=1
                     /product="E1 and E2 glycoproteins"
                     /protein_id="AAF26710.1"
                     /db_xref="GI:6716710"
                     /translation="MASTTPITMEDLQKALEAQSRALRAELAAGASQPRRPRPPRQRD
                     SSTSGDDSGRDSGGPRRRRGNRGRGQRKDWSRAPPPPEERQEGRSQTPAPKPSRAPPQ
                     QPQPPRMQTGRGGSAPRPELGPPTNPFQAAVARGLRPPLHDPDTEAPTEACVTSWLWS
                     EGEGAVFYRVDLHFTNLGTPPLDEDGRWDPALMYNPCGPEPPAHVVRAYNQPAGDVRG
                     VWGKGERTYAEQDFRVGGTRWHRLLRMPVRGLDGDSAPLPPHTTERIETRSARHPWRI
                     RFGAPQAFLAGLLLAAVAVGTARAGLQPRVDMAAPPTPPQPPRAHGQHYGHHHHQLPF
                     LGHDGHHGGTLRVGQHHRNASDVLPGHWLQGGWGCYNLSDWHQGTHVCHTKHMDFWCV
                     EHDRPPPATPTPLTTAANSITAATPATAPAPCHAGLNDSCGGFLSGCGPMRLRHGADT
                     RCGRLICGLSTTAQYPPTRFGCAMRWGLPPWELVVLTARPEDGWTCRGVPAHPGTRCP
                     ELVSPMGRATCSPASALWLATANALSLDHALAAFVLLVPWVLIFMVCRRACRRRGAAA
                     ALTAVVLQGYNPPAYGEEAFTYLCTAPGCATQTPVPVRLACVRFESKIVDGGCFAPWD
                     LEATGACICEIPTDVSCEGLGAWVPTAPCARIWNGTQRACTFWAVNAYSSGGYAQLAS
                     YFNPGGSYYKQYHPTACEVEPAFGHSDAACWGFPTDTVMSVFALASYVQHPHKTVRVK
                     FHTETRTVWQLSVAGVSCDVTTEHPFCNTPHGQLEVQVPPDPGDMVEYIMNYTGNQQS
                     RWGLGSPNCHGPDWASPVCQRHSPDCSRLVGATPERPRLRLVDADDPLLRTAPGPGEV
                     WVTPVIGSQARKCGLHIRAGPYGHATVEMPEWIHAHTTSDPWHPPGPLGLKFKTVRPV
                     TLPRALAPPRNVRVTGCYQCGTPALVEGLAPGGGNCHLTVNGEDVGAFPPGKFVTAAL
                     LNTPPPYQVSCGGESDRASARVIDPAAQSFTGVVYGTHTTAVSETRQTWAEWAAAHWW
                     QLTLGAICALPLAGLLACCAKCLYYLRGAIAPR"
BASE COUNT     1467 a   3765 c   3006 g   1524 t
ORIGIN


  Query Match          2.8%;  Score 37.2;  DB 14;  Length 9762;
  Best Local Similarity   45.7%;  Pred. No. 2.2e+02;
  Matches  123;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy      55 gtcatttatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggt 114
           | ||  | || || | ||| |         | |  |  |     ||| ||  . |  |
Db    3944 GGCACCTGTGCCGCGACCGACGAGGGGCTGGCCCAGGCGTACTACGACGACCTCGAGGTG 4003

Qy     115 atgcgcccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgttt 174
             |||| |  |    | | || ||   | |  ||| ||  |  || ||||   |
Db    4004 CGCCGCCTCGGGGATGACGCCATGGCCCGGGCGGCCCTCGCATCAGTCCAACGCCCTCGC 4063

Qy     175 gaagacaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgcc 234
           |||  |    |  ||   | ||   |  ||  | |   | |||||  | |  ||  |
Db    4064 AAAGGCCCTTACAATATCAGGGTATGGAACATGGCCGCAGGCGCTGGCAAGACCACCCGC 4123

Qy     235 atccatcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgac 294
           ||||     | |   | ||||| |   |  |  |||       ||  ||    |    ||
Db    4124 ATCCTCGCTGCCTTCACGCGCGAAGACCTCTACGTCTGCCCCACCAATGCGCTCCTGCAC 4183

Qy     295 gaaatcgagttcgaacgctacgcgcccga 323
           || ||| ||  | |||  |  ||||| |||
Db    4184 GAGATCCAGGCCAAACTCCGCGCGCGCGA 4212


RESULT  41
RUBNOPR
LOCUS       RUBNOPR                 9762 bp ss-RNA     linear   VRL 14-AUG-2001
DEFINITION  Rubella virus nonstructural protein and virion protein genes,
            complete cds.
ACCESSION   L78917
VERSION     L78917.1  GI:2529630
KEYWORDS    .
SOURCE      Rubella virus.
  ORGANISM  Rubella virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
            Rubivirus.
REFERENCE   1  (bases 1 to 9762)
  AUTHORS   Pugachev,K.V., Abernathy,E.S. and Frey,T.K.
  TITLE     Genomic sequence of the RA27/3 vaccine strain of rubella virus
  JOURNAL   Arch. Virol. 142 (6), 1165-1180 (1997)
  MEDLINE   97372951
   PUBMED   9229006
COMMENT     GSDB:S:110169
FEATURES             Location/Qualifiers
     source          1. .9762
                     /organism="Rubella virus"
                     /isolate="RA 27/3"
                     /db_xref="taxon:11041"
     mat_peptide     41. .3943
                     /product="P150- N-terminal nonstructural protein from
                     nonstructural protein precursor; putative"
     CDS             41. .6391
                     /note="nonstructural protein open reading frame
                     (translated into a precursor that is then cleaved into two
                     individual components); putative"
                     /codon_start=1
                     /product="nonstructural protein"
                     /protein_id="AAB81187.1"
                     /db_xref="GI:2529631"
                     /translation="MERLLDEVLAPGGPYNLTVGSWVRDHVRSIVEGAWEVRDVVSAA
                     QKRAIVAVIPRPVFTQMQVSDHPALHAISRYTRRHWIEWGPKEALHVLIDPSPGLLRE
                     VARVERRWVALCLHRTARKLATALAETASEAWHADYVCALRGAPSGPFYVHPEDVPHG
                     GRAVADRCLLYYTPMQMCELMRTIDATLLVAVDLWPVALAAHVGDDWDDLGIAWHLDH
                     DGGCPADCRGAGAGPTPGYTRPCTTRIYQVLPDTAHPGRLYRCGPRLWTRDCAVAELS
                     WEVAQHCGHQARVRAVRCTLPIRHVRSLQPSARVRLPDLVHLAEVGRWRWFSLPRPVF
                     QRMLSYCKTLSPDAYYSERVFKFKNALSHSITLAGNVLQEGWKGTCAEEDALCAYVAF
                     RAWQSNARLAGIMKSAKRCAADSLSVAGWLDTIWGAIKRFFGSVPLAERMEEWEQDAA
                     VAAFDRGPLEDGGRHLDTVQPPKSPPRPEIAATWIVHAASADRHCACAPRCDVPRERP
                     SAPAGPPDDEALIPPWLFAEHRALRCREWDFEVLRARADTAAAPAPLAPRPARYPTVL
                     YRHPAHHGPWLTLDEPGEADAALVLCDPLGQPLRGPERHFAAGAHMCAQARGLQAFVR
                     VVPPPERPWADGGARAWAKFFRGCAWAQRLLGEPAVMHLPYTDGDVPQLIALALRTLA
                     QQGAALALSVRDLPGGAAFDANAVTAAVRAGPGQSAATSSPPGDPPPPRCARRSQRHS
                     DARGTPPPAPARDPPPPAPSPPAPPRAGDPVPPTSAGPADRARDAELEVAYEPSGPPT
                     STKADPDSDIVESYARAAGPVHLRVRDIMDPPPGCKVVVNAANEGLLAGSGVCGAIFA
                     NATAALAADCRRLAPCPTGEAVATPGHGCGYTHIIHAVAPRRPRDPAALEEGEALLER
                     AYRSIVALAAARRWARVACPLLGAGVYGWSAAESLRAALAATRTEPAERVSLHICHPD
                     RATLTHASVLVGAGLAARRVSPPPTEPLASCPAGDPGRPAQRSASPPATPLGDATAPE
                     PRGCQGCELCRYTRVTNDRAYVNLWLERDRGATSWAMRIPEVVVYGPEHLATHFPLNH
                     YSVLKPAEVRPPRGMCGSDMWRCRGWQGVPQVRCTPSNAHAALCRTGVPPRVSTRGGE
                     LDPNTCWLRAAANVAQAARACGAYTSAGCPRCAYGRALSEARTHKDFAALSQRWSASH
                     ADASSDGTGDPLDPLMETVGCACSRVWVGSEHEAPPDHLLVSLHRAPNGPWGVVLEVR
                     ARPEGGNPTGHFVCAVGGGPRRVSDRPHLWLAVPLSRGGGTCAATDEGLAQAYYDDLE
                     VRRLGDDAMARAALASVQRPRKGPYNIRVWNMAAGAGKTTRILAAFTREDLYVCPTNA
                     LLHEIQAKLRARDIEIKNAATYERALTKPLAAYRRIYIDEAFTLGGEYCAFVASQTTA
                     EVICVGDRDQCGPHYANNCRTPVPDRWPTERSRHTWRFPDCWAARLRAGLDYDIEGER
                     TGTFACNLWDGRQVDLHLAFSRETVRRLHEAGIRAYTVREAQGMSVGTACIHVGRDGT
                     DVALALTRDLAIVSLTRASDALYLHELEDGSLRAAGLSAFLDAGALAELKEVPAGIDR
                     VVAVEQAPPPLPPADGIPEAQDVPPFCPRTLEELVFGRAGHPHYADLNRVTEGEREVR
                     YMRISRHLLNKNHTEMPGTERVLSAVCAVRRYRAGEDGSTLRTAVARQHPRPFRQIPP
                     PRVTAGVAQEWRMTYLRERIDLTDVYTQMGVAARELTDRYARRYPEIFAGMCTAQSLS
                     VPAFLKATLKCVDAALGPRDTEDCHAAQGKAGLEIRAWAKEWVQVMSPHFRAIQKIIM
                     RALRPQFLVAAGHTEPEVDAWWQAHYTTNAIEVDFTEFDMNQTLATRDVELEISAALL
                     GLPCAEDYRALRAGSYCTLRELGSTETGCERTSGEPATLLHNTTVAMCMAMRMVPKGV
                     RWAGIFQGDDMVIFLPEGARSAALKWTPAEVGLFGFHIPVKHVSTPTPSFCGHVGTAA
                     GLFHDVMHQAIKVLCRRFDPDVLEEQQVALLDRLRGVYAALPDTVAANAAYYDYSAER
                     VLAIVRELTAYARGRGLDHPATIGALEEIQTPYARANLHDAD"
     mat_peptide     3947. .6388
                     /product="P90- C-terminal nonstructural protein from
                     nonstructural protein precursor; putative"
     misc_RNA        6436. .9762
                     /note="subgenomic RNA for translation of structural
                     protein ORF; putative"
     mat_peptide     6512. .7411
```

```
                     /note="putative"
                     /product="capsid protein"
     CDS             6512. .9703
                     /note="structural protein open reading frame; putative"
                     /codon_start=1
                     /product="virion protein"
                     /protein_id="AAB81188.1"
                     /db_xref="GI:2529632"
                     /translation="MASTTPITMEDLQKALETQSRALRAELAAGASQSRRPRPPRQRD
                     SSTTGDDSGRDSGGPRRRRGNRGRGQRRDWSRAPPPPEERQETRSQTPAPKPSRAPPQ
                     QPQPPPRMQTGRGGSAPRPELGPPTNPFQAAVARGLRPPLHDPDTEAPTEACVTSWLWS
                     EGEGAVFYRVDLHFTNLGTPPLDEDGRWDPALMYNPCGPEPPAHVVRAYNQPAGDVRG
                     VWGKGERTYAEQDFRVGGTRWHRLLRMPVRGLDGDSAPLPPHTTERIETRSARHPWRI
                     RFGAPQAFLAGLLLAAVAVGTARAGLQPRADMAAPPTLPQPPRAHGQHYGHHHHQLPF
                     LGHDGHHGGTLRVGQHHRNASDVLPGHWLQGGWGCYNLSDWHQGTHVCHTKHMDFWCV
                     EHDRPPPATPTPLTTAANSTTAATPATAPAPCHAGLNDSCGGFLSGCGPMRLRHGADT
                     RCGRLICGLSTTAQYPPTRFGCAMRWGLPPWELVVLTARPEDGWTCRGVPAHPGTRCP
                     ELVSPMGRATCSPASALWLATANALSLDHALAAFVLLVPWVLIFMVCRRTCRRRGAAA
                     ALTAVVLQGYNPPAYGEEAFTYLCTAPGCATQAPVPVRLAGVRFESKIVDGGCFAPWD
                     LEATGACICEIPTDVSCEGLGAWVPTAPCARIWNGTQRACTFWAVNAYSSGGYAQLAS
                     YFNPGGSYYKQYHPTACEVEPAFGHSDAACWGFPTDTVMSVFALASYVQHPHKTVRVK
                     FHTETRTVWQLSVAGVSCNVTTEHPFCNTPHGQLEVQVPPDPGDLVEYIMNHTGNQQS
                     RWGLGSPNCHGPDWASPVCQRHSPDCSRLVGATPERPRLRLVDADDPLLRTAPGPGEV
                     WVTPVIGSQARKCGLHIRAGPYGHATVEMPEWIHAHTTSDPWHPPGPLGLKFKTVRPV
                     ALPRTLAPPRNVRVTGCYQCGTPALVEGLAPGGGNCHLTVNGEDLGAFPPGKFVTAAL
                     LNTPPPYQVSCGGESDRASARVIDPAAQSFTGVVYGTHTTAVSETRQTWAEWAAAHWW
                     QLTLGAICALLLAGLLACCAKCLYYLRGAIAPR"
     mat_peptide     7412. .8257
                     /note="putative"
                     /product="E2 envelope glycoprotein"
     mat_peptide     8258. .9700
                     /note="putative"
                     /product="E1 envelope glycoprotein"
BASE COUNT     1458 a   3776 c   3010 g   1518 t
ORIGIN
```

```
  Query Match             2.8%;  Score 37.2;  DB 14;  Length 9762;
  Best Local Similarity   45.7%;  Pred. No. 2.2e+02;
  Matches  123;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy      55 gtcatttatgacgggcccgtcattaccgaagtcgcgttgcttggcgaagaatatgccggt 114
           | ||  | || || | ||| |        | |  |  |     ||| ||    |  |
Db    3944 GGCACCTGTGCCGCGACCGACGAGGGGCTGGCCCAGGCGTACTACGACGACCTCGAGGTG 4003

Qy     115 atgcgccccttngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaagtgctgtttt 174
           |||| |  |    | | || ||   | |  ||| ||  |  || ||||   |
Db    4004 CGCCGCCTCGGGGATGACGCCATGGCCCGGGCGGCCCTCGCATCAGTCCAACGCCCTCGC 4063

Qy     175 gaagacaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaatcgccgcc 234
           ||| |    |  ||    | ||    |   |  ||  | |   | ||||| | | ||  |
Db    4064 AAAGGCCCTTACAATATCAGGGTATGGAACATGGCCGCAGGCGCTGGCAAGACCACCCGC 4123

Qy     235 atccatcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgac 294
           ||||     | |   | |||||  |   |  |  |||       ||  ||    |    ||
Db    4124 ATCCTCGCTGCCTTCACGCGCGAAGACCTTTACGTCTGCCCCACCAATGCGCTCCTGCAC 4183

Qy     295 gaaatcgagttcgaacgctacgcgcccga 323
           || ||| ||  | ||| |  ||||| |||
Db    4184 GAGATCCAGGCCAAACTCCGCGCGCGCGA 4212
```

```
RESULT  42
AC016780/c
LOCUS       AC016780              112721 bp    DNA     linear   PLN 14-JUL-2000
DEFINITION  Genomic Sequence For Oryza sativa, Niponbare Strain, Clone
            OSJNBa0061K21 From Chromosome 10, complete sequence.
ACCESSION   AC016780
VERSION     AC016780.6  GI:9186821
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 112721)
  AUTHORS   Spiegel,L.A., Nascimento,L.U., de la Bastide,M., Vil,D.M.,
            Preston,R.R., Huang,E.N., See,L.H., Shah,R.S., O'Shaughnessy,A.,
            Rodriguez,M.A., Shekher,M., Kirchoff,K.A., Baker,J.P., Schutz,K.,
            Dedhia,N.N. and McCombie,W.R.
  TITLE     Genomic Sequence For Oryza sativa, Niponbare Strain, Clone
            OSJNBa0061K21 From Chromosome 10, Complete Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112721)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1999) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
REFERENCE   3  (bases 1 to 112721)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     On Jul 14, 2000 this sequence version replaced gi:7527768.
FEATURES             Location/Qualifiers
     source          1. .112721
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="10"
                     /clone="OSJNBa0061K21"
                     /clone_lib="Hind III"
     misc_feature    44625. .44670
                     /note="The assembly is single stranded with single
                     chemistry."
     misc_feature    50600. .50630
                     /note="The assembly is single stranded with single
                     chemistry."
     misc_feature    53490. .53680
                     /note="The assembly is single stranded with single
                     chemistry."
     misc_feature    53705. .53770
                     /note="The assembly is single stranded with single
                     chemistry."
     misc_feature    69610. .69690
                     /note="This area is designated as a low coverage area."
     misc_feature    91500. .91540
                     /note="The assembly is designated low coverage."
     misc_feature    108580. .108680
                     /note="The assembly is single stranded with single
                     chemistry."
BASE COUNT    32366 a  24371 c  23231 g  32753 t
ORIGIN
```

```
  Query Match             2.8%;  Score 37.2;  DB 8;  Length 112721;
  Best Local Similarity   47.5%;  Pred. No. 2.3e+02;
  Matches  105;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy     107 atgccggtatgcgcccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaag 166
           | | ||   |  ||  |  | ||    |||||  |||||  | ||||     | |   |
Db  109836 ACGTCGTCGTCGGCGACGTGGTGGTGCTCAAGATCGGCGACGTCGTCCCCGCCGACGGCG 109777

Qy     167 tgctgtttgaagacaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaa 226
           || |  | || | | |   |   | || ||  | ||  | |  | | |         |
Db  109776 TGTTCCTGGACGGCCACGCGCTGCAGGTCGACGAGTCGAGCATGACCGGCGAGCCGCACC 109717

Qy     227 tcgccgccatccatcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaag 286
           || ||   || |   || | | |  | | | ||    || | ||| |  |  || || |
Db  109716 CCGTCGAGGTCGACGCCGTCAAGAGCCCCTTCCTCGCCTCCGGCGTCAAGGTGGTCGACG 109657

Qy     287 gcaacgacgaaatcgagttcgaacgctacgcgcccgaagcg 327
           || ||| | | || | | ||      |  || | | ||  ||
Db  109656 GCTACGGCAAGATGGTGGTCACCGCCGTCGGGACGGACACG 109616
```

```
RESULT 43
AC027658/c
LOCUS       AC027658               142737 bp    DNA     linear   PLN 02-APR-2000
DEFINITION  Oryza sativa subsp. japonica BAC nbxb0006I13, chromosome 10,
            complete sequence.
ACCESSION   AC027658
VERSION     AC027658.1  GI:7363409
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 142737)
  AUTHORS   Llaca,V., Young,S., Kavchok,S., Clark,T., Choi,V.,
            Nevill-Manning,C. and Messing,J.
  TITLE     Rice Chromosome 10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142737)
  AUTHORS   Llaca,V., Lou,A., Young,S., Charydczac,G., Choi,V.,
            Nevill-Manning,C. and Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) The Plant Genome Initiative at Rutgers -
            Waksman Institute, Rutgers University, 190 Frelinghuysen Road,
            Piscataway, NJ 08873
REFERENCE   3  (bases 1 to 142737)
  AUTHORS   Llaca,V., Young,S., Kavchok,S., Clark,T., Charydczac,G., Choi,V.,
            Nevill-Manning,C. and Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2000) The Plant Genome Initiative at Rutgers -
            Waksman Institute, Rutgers University, 190 Frelinghuysen Road,
            Piscataway, NJ 08873
  REMARK    Chromosome 10
FEATURES             Location/Qualifiers
     source          1. .142737
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="10"
                     /clone="nbxb0006I13"
BASE COUNT    42049 a  28618 c  30051 g  42019 t
ORIGIN


  Query Match             2.8%;  Score 37.2;  DB 8;  Length 142737;
  Best Local Similarity   47.5%;  Pred. No. 2.3e+02;
  Matches  105;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy     107 atgccggtatgcgccccctngatgaaagtcaaggaaggcgatgccgtcaaaaaaggccaag 166
           | | ||   |  ||  |  | ||    |||||  ||||| | ||||     | |   |
Db   42381 ACGTCGTCGTCGGCGACGTGGTGGTGCTCAAGATCGGCGACGTCGTCCCCGCCGACGGCG 42322

Qy     167 tgctgtttgaagacaaaaagnatccgggcgtggtgtttaccgcgccngtttcaggcaaaa 226
           || |  | || | | |   |   | || ||  | ||  | |  | | |         |
Db   42321 TGTTCCTGGACGGCCACGCGCTGCAGGTCGACGAGTCGAGCATGACCGGCGAGCCGCACC 42262

Qy     227 tcgccgccatccatcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaag 286
           || ||   || |   || | | |  | | | ||    || | ||| |  |  || || |
Db   42261 CCGTCGAGGTCGACGCCGTCAAGAGCCCCTTCCTCGCCTCCGGCGTCAAGGTGGTCGACG 42202

Qy     287 gcaacgacgaaatcgagttcgaacgctacgcgcccgaagcg 327
           || ||| | | || | | ||     |  || | | ||  ||
Db   42201 GCTACGGCAAGATGGTGGTCACCGCCGTCGGGACGGACACG 42161


RESULT 44
NMA5Z2491
LOCUS       NMA5Z2491              329861 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 5/7.
ACCESSION   AL162756 AL157959
VERSION     AL162756.2  GI:7380091
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 329861)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 329861)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .329861
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            194. .565
                     /gene="NMA1439"
     CDS             194. .565
                     /gene="NMA1439"
                     /note="NMA1439, possible lipoprotein, len: 123 aa;
                     unknown, contains a probable N-terminal signal sequence
                     and an appropriately positioned PS00013 Prokaryotic
                     membrane lipoprotein lipid attachment site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="CAB84675.1"
                     /db_xref="GI:7380092"
                     /db_xref="SPTREMBL:Q9JU92"
                     /translation="MNKTLSILPVAILLGGCAAGGGNTFGSLDGGTGMGGSIVKMAVE
                     SQCRAELNKRSEWRLTALAMSAEKQAEWENKICACVAQEAPNQLTGNDVMQMLDPSTR
                     NQALAALTAKTVSACFKHLYR"
     misc_feature    212. .244
                     /gene="NMA1439"
                     /note="PS00013 Prokaryotic membrane lipoprotein lipid
                     attachment site"
     misc_feature    complement(571. .580)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    592. .601
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            638. .2047
                     /gene="thrC"
     CDS             638. .2047
                     /gene="thrC"
                     /EC_number="4.2.99.2"
                     /note="NMA1440, thrC, probable threonine synthase, len:
                     485 aa; similar to many e.g. SW:THRC_METGL (EMBL:D14071),
                     thrC, Methylobacillus glycogenes threonine synthase (EC
                     4.2.99.2) (475 aa), fasta scores; E(): 0, 61.9% identity
                     in 475 aa overlap. Contains Pfam match to entry PF00291
                     S_T_dehydratase, Pyridoxal-phosphate dependent enzymes"
                     /codon_start=1
                     /transl_table=11
                     /product="putative threonine synthase"
                     /protein_id="CAB84676.1"
```

```
                     /db_xref="GI:7380093"
                     /db_xref="SPTREMBL:Q9JU91"
                     /translation="MKYISTRGETAHKPFSEVLLMGLAPDGGLMLPEHYPQIGREALD
                     KWRGLAYPELAFEIMRLFVTDIPEDDLRDILNRTYTEAAFGTKEITPVRTLSDGIKIQ
                     ALSNGPTLAFKDMAMQFLGNAFEYVLNKEGKKLNILGATSGDTGSAAEYALRGKKGVN
                     VFMLSPDGKMSAFQRAQMYSLQDGNIHNIAVKGMFDDCQDIVKAVQNDAAFKEKYHIG
                     TVNSINWGRIVAQVVYYFAGYFKATQSNDEQVSFCVPSGNFGNVCAGHIAKQMGLPVR
                     RLIVATNENDVLDEFFKTGAYRPRNSAHTYVTSSPSMDISKASNFERFVFDLMDRDPQ
                     EINTLWAEVAAGKGFNLRFALDKVGGKYGFTSGKSTHADRLATIRQVYEQDQELIDPH
                     TADGVKVAREVREERETVVCLETALAAKFDATIREAVGDVAIPRPAALEGLENLPQRV
                     QTVPNSADAVKGIIEQTLA"
misc_feature         875. .1780
                     /gene="thrC"
                     /note="Pfam match to entry PF00291 S_T_dehydratase,
                     Pyridoxal-phosphate dependent enzymes, score 231.80,
                     E-value 1e-65"
misc_feature         complement(922. .931)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
gene                 2094. .2888
                     /gene="NMA1441"
CDS                  2094. .2888
                     /gene="NMA1441"
                     /note="NMA1441, len: 264 aa; unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA1441"
                     /protein_id="CAB84677.1"
                     /db_xref="GI:7380094"
                     /db_xref="SPTREMBL:Q9JU90"
                     /translation="MTVPMQETHFSILLDELAAKQEAAIAPHLLADGTKVWIRKAGRH
                     NARWRYALLGMVARYLKLGVLKPVPSLGGEPAIATESKRLYELRAAGIAVPELLAVRK
                     NALMFGNLEGIPLDTQIRQEAEAGKADAWLAGLEAIARVHKKQQFLSQAFARNMMSDG
                     KNISFLDFEDDPSEVLTIAQCQARDWLCYIHSTALILKNGGLLEAAAEKWGGVLSDQP
                     AEIQKLIAATVKPILPIRRLEHPRWGRDALRLAASISLISLADMPP"
stem_loop            2904. .2937
                     /note="Stem loop containing DNA uptake sequences: acgat
                     gccgtctgaa gccc ttcagacggc atata"
misc_feature         2909. .2918
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
misc_feature         complement(2923. .2932)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
RBS                  3065. .3068
gene                 3076. .3852
                     /gene="fpr"
CDS                  3076. .3852
                     /gene="fpr"
                     /EC_number="1.18.1.2"
                     /note="NMA1442, fpr, probable ferredoxin--NADP reductase,
                     len: 258 aa; similar to many e.g. SW:FENR_AZOVI
                     (EMBL:L36319), fpr, Azotobacter vinelandii
                     ferredoxin--NADP reductase (EC 1.18.1.2) (257 aa), fasta
                     scores; E(): 0, 69.8% identity in 255 aa overlap. Similar
                     to NMA1664, fasta scores; E(): 1.3e-30, 34.6% identity in
                     257 aa overlap. Contains Pfam match to entry PF00175
                     oxidored_fad, Oxidoreductase FAD/NAD-binding domain"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ferredoxin--NADP reductase"
                     /protein_id="CAB84678.1"
                     /db_xref="GI:7380095"
                     /db_xref="SPTREMBL:Q9JRE3"
                     /translation="MAAFNTQKVLSVHHWTDAYFTFTCTRDESLRFENGQFVMVGLMV
                     DGKPLMRAYSVASANWEEHLEFFSIKVQDGPLTSRLQHLKVGDDVLISKKPTGTLVAG
                     DLNPGKHLYLLLSTGTGIAPFLSITKDPEIYEQFEKIILVHGVRYKKDLAYYDRFTKEL
                     PEHEYLGDLVKEKLIYYPIVSREEFEHHGRLTDLMVSGKLFEDIGLPKINPQDDRAML
                     CGSPAMLKDTCKVLDDFGLTVSPKTGVRGDYLIERAFVDQ"
misc_feature         3394. .3768
                     /gene="fpr"
                     /note="Pfam match to entry PF00175 oxidored_fad,
                     Oxidoreductase FAD/NAD-binding domain, score 20.00,
                     E-value 8.4e-05"
misc_feature         3866. .3875
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
repeat_unit          3964. .3983
                     /note=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
                     /label=dRS3
repeat_unit          complement(3984. .4038)
                     /note="RS103"
                     /label=RS103
repeat_unit          4039. .4058
                     /note=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
                     /label=dRS3
gene                 4159. .4320
                     /gene="NMA1443"
CDS                  4159. .4320
                     /gene="NMA1443"
                     /note="NMA1443, len: 53 aa; unknown, lies within a region
                     of unusually low GC content"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA1443"
                     /protein_id="CAB84679.1"
                     /db_xref="GI:7380096"
                     /db_xref="SPTREMBL:Q9JU89"
                     /translation="MIYKKNCPNKKCYAENKMKRMVFIENGCSATVIHRTGIWKKIIH
                     IDIIIPFIA"
gene                 complement(4509. .6980)
                     /gene="NMA1444"
CDS                  complement(4509. .6980)
                     /gene="NMA1444"
                     /note="NMA1444, probable P-type cation-transporting
                     ATPase, len: 823 aa; similar to many e.g. SW:COPA_ENTHR
                     (EMBL:L13292), copA, Enterococcus hirae
                     copper/potassium-transporting ATPase A (EC 3.6.1.36) (727
                     aa), fasta scores; E(): 0, 34.0% identity in 744 aa
                     overlap. Similar to NMA1539, fasta scores; E(): 0, 34.4%
                     identity in 735 aa overlap. Contains hydrophobic, probable
                     membrane-spanning regions. Contains two Pfam matches to
                     entry PF00122 E1-E2_ATPase, E1-E2 ATPases and PS00154
                     E1-E2 ATPases phosphorylation site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative P-type cation-transporting ATPase"
                     /protein_id="CAB84680.1"
                     /db_xref="GI:7380097"
                     /db_xref="SPTREMBL:Q9JU88"
                     /translation="MKKTCFHCGLDVPENLHLTVRYENEDRETCCAGCQAVAQSIIDA

  Query Match              2.8%;  Score 37.2;  DB 1;  Length 329861;
  Best Local Similarity   48.2%;  Pred. No. 2.4e+02;
  Matches   93;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     237 ccatcgcggcgaaaagcgcgtacttcagtcggtcgtgattgccgttgaaggcaacgacga 296
           | | | |  ||  | |  ||  |  | | | | | ||     |   |||  |  || |
Db 161098 CAAACACATCGGCACGTTCGCCCACCTGACCGCCCTGCGCCGCACCGAAACCGCCGGCTT 161157

Qy     297 aatcgagttcgaacgctacgcgcccgaagcgttggcaaacttaagcggcgangaantnng 356
           | |     | || || || ||| |||||| ||||||||||    ||     |||
Db 161158 TACCATCGCCCAAAGCCACACGCTCGAAGCCTTGGCAAACTTGGACGAAACAGAACGCGA 161217

Qy     357 nngcaatctgatccaatccggtttgtggactgcgctgcgtanccgtccgttcagcaaaat 416
             ||   ||| | |  | ||   | | |  ||| |    |  ||   |  |    ||
Db 161218 CAGCCTGCTGCTGCCCTGCGACGTATTGGTTGCCCATTTTCCCCAAACCGTTTTAAACGA 161277

Qy     417 ccctgccgtcgat 429
             ||||||| ||
Db 161278 TTATGCCGTCCAT 161290


RESULT  45
AX019436/c
LOCUS       AX019436                763 bp    DNA     linear   PAT 07-SEP-2000
```

```
DEFINITION  Sequence 13 from Patent WO9940109.
ACCESSION   AX019436
VERSION     AX019436.1  GI:10043375
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 763)
  AUTHORS   Welters,P., Stiefel,T., Voelter,W. and Morris,P.
  TITLE     Recombinant mistletoe lectines
  JOURNAL   Patent: WO 9940109-A 13 12-AUG-1999;
            WELTERS PETER (DE); STIEFEL THOMAS (DE); BIOSYN ARZNEIMITTEL GMBH
            (DE); VOELTER WOLFGANG (DE); MORRIS PETER (GB)
FEATURES             Location/Qualifiers
     source          1. .763
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
     misc_feature    319
BASE COUNT      163 a    215 c    175 g    170 t     40 others
ORIGIN


  Query Match             2.8%;  Score 37;  DB 6;  Length 763;
  Best Local Similarity   47.6%;  Pred. No. 2.4e+02;
  Matches   79;  Conservative    8;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     871 gaattggttgacgcagacaaccgcgtgatttccggttcggtattgaacggcgcgattaca 930
           || | || ||  || |   |    |||  || | |  | |  | |  || ||||   | |
Db     321 GANTCGGGTGGTGCCGGTGAAGAGGTGCGTTTCCGCGCCGCGTGGTGCGTCGCGCAAAAA 262

Qy     931 caaggcgcgcacgattatttgggacgctaccacaatcagatttccgttatcgaagaaggc 990
             |||   |  ||  |  |||| | || || ||   :::|||       |||| |  |||
Db     261 GTAGGATTGGTCGCCTGCTTGGTAAGCCACGACGTAMRSATTGGTAACGTCGATGGCGGC 202

Qy     991 cgcagcaaagagctgttcggctgggttgcgccgcagccggacaaat 1036
           || : :  ||  :::  |      | || ||  || |  |||||||
Db     201 CGTRAYCGAGTCYYSCCCCTGGTTGGTGAGCTCCACCAAGACAAAT 156


Search completed: June 30, 2002, 07:14:05
Job time: 50263 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 07:14:05 ; Search time 27321.5 Seconds
(without alignments)
1029.420 Million cell updates/sec

Title: US-09-303-518D-131
Perfect score: 1344
Sequence: 1 atgattaaaatcaaaaaagg.....ccattgagaaggaaggctga 1344

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1344	100.0	1344	6	A96098	A96098 Sequence 13
	2	1212.8	90.2	1344	6	A96092	A96092 Sequence 12
c	3	1212.8	90.2	12273	1	AE002412	AE002412 Neisseria
c	4	1212.8	90.2	349980	6	AX044029	AX044029 Sequence
c	5	1179.2	87.7	311321	1	NMA3Z2491	AL162754 Neisseria
	6	1165.4	86.7	1344	6	A96094	A96094 Sequence 12
	7	574.4	42.7	10563	1	AE006171	AE006171 Pasteurel
	8	566.2	42.1	1558	1	APU24492	U24492 Actinobacil
	9	550.2	40.9	13299	1	U32702	U32702 Haemophilus
	10	518.4	38.6	5916	1	AF117331	AF117331 Vibrio ch
c	11	518.4	38.6	13382	1	AE004300	AE004300 Vibrio ch
	12	513.6	38.2	6048	1	AF165980	AF165980 Vibrio ha
	13	488	36.3	5078	1	VANQRBOL	Z37111 V.alginolyt
	14	483.2	36.0	6908	1	AB008030	AB008030 Vibrio al
c	15	475.6	35.4	16903	1	AE004724	AE004724 Pseudomon
	16	433.6	32.3	474	6	A96090	A96090 Sequence 12
c	17	348.8	26.0	96109	6	AX067460	AX067460 Sequence
	18	126.4	9.4	1356	1	AF144076	AF144076 Porphyrom
c	19	105.4	7.8	4926	1	HIU20229	U20229 Haemophilus
c	20	42.4	3.2	36200	1	BACGNTZA	D78193 Bacillus su
	21	42.4	3.2	151329	2	AP003456	AP003456 Oryza sat
c	22	42.4	3.2	215534	1	BSUB0021	Z99124 Bacillus su
c	23	40.2	3.0	125020	9	AF429315	AF429315 Homo sapi
c	24	40	3.0	37000	14	BHT1UL	Z78205 Bovine herp
c	25	40	3.0	135301	14	BHV1CGEN	AJ004801 Bovine he
	26	39.8	3.0	216050	1	AL646076	AL646076 Ralstonia
	27	39	2.9	125020	9	AF429315	AF429315 Homo sapi
	28	38.8	2.9	8918	1	AB047272	AB047272 Pseudomon
	29	38.8	2.9	123580	1	AF263912	AF263912 Streptomy
	30	38.8	2.9	125401	6	AX211739	AX211739 Sequence
	31	38.4	2.9	10500	1	AE005062	AE005062 Halobacte
	32	38.2	2.8	3000	1	AF220510	AF220510 Azoarcus
c	33	38.2	2.8	9980	1	AE004915	AE004915 Pseudomon
	34	38.2	2.8	38734	1	SC10A5	AL021529 Streptomy
c	35	38	2.8	1641	1	AB015974	AB015974 Pseudomon
c	36	37.8	2.8	340900	1	SME591791	AL591791 Sinorhizo
	37	37.6	2.8	760	4	AY011741	AY011741 Canis fam
	38	37.2	2.8	1731	1	KPU81143	U81143 Klebsiella
c	39	37.2	2.8	4130	8	AF190427	AF190427 Colletotr
c	40	37.2	2.8	7218	6	I66494	I66494 Sequence 14
	41	37.2	2.8	213050	1	AL646079	AL646079 Ralstonia
c	42	37	2.8	10029	1	AE008180	AE008180 Agrobacte
c	43	37	2.8	10359	1	AE009214	AE009214 Agrobacte
c	44	36.8	2.7	10514	1	AE002070	AE002070 Deinococc
	45	36.4	2.7	3855	13	M23555	M23555 Figure 2. N
	46	36.4	2.7	3867	1	AY033407	AY033407 Myxococcu
	47	36.4	2.7	189050	1	AL646069	AL646069 Ralstonia
c	48	36.2	2.7	30132	1	SC10G8	AL158057 Streptomy
	49	36.2	2.7	110000	2	LMFLCHR32_04	Continuation (5 of
c	50	36.2	2.7	174301	2	AP003523	AP003523 Oryza sat
	51	36	2.7	11526	1	AF228498	AF228498 Escherich
	52	35.8	2.7	16897	1	AE006941	AE006941 Mycobacte
	53	35.8	2.7	35019	1	MTY13E10	Z95324 Mycobacteri
	54	35.8	2.7	154330	2	AC104429	AC104429 Oryza sat
c	55	35.8	2.7	157912	9	AL133387	AL133387 Human DNA
	56	35.6	2.6	632	3	AF240033	AF240033 Asterias
	57	35.6	2.6	960	1	PSU83882	U83882 Pseudomonas
	58	35.6	2.6	1462	1	PSECATEX	L10655 Sphingomona
	59	35.6	2.6	1918	1	AB008137	AB008137 Enterobac
	60	35.6	2.6	1930	1	AB008140	AB008140 Enterobac
	61	35.6	2.6	4004	1	SSPZ84817	Z84817 Sphingomona
c	62	35.6	2.6	12129	1	AE007223	AE007223 Sinorhizo
c	63	35.6	2.6	17205	1	AE007185	AE007185 Mycobacte
c	64	35.6	2.6	24268	1	AB070947	AB070947 Streptomy
c	65	35.6	2.6	30352	1	MTCY409	Z97188 Mycobacteri
	66	35.6	2.6	41321	1	MSGY409	AD000017 Mycobacte
	67	35.4	2.6	296	6	AX260544	AX260544 Sequence
	68	35.4	2.6	10732	6	E32986	E32986 Gene encodi
	69	35.2	2.6	3955	14	NPOEGT	D17353 Orgyia pseu
c	70	35.2	2.6	10234	1	AE009378	AE009378 Agrobacte
	71	35.2	2.6	10647	1	AE008237	AE008237 Agrobacte
	72	35.2	2.6	131995	14	OPU75930	U75930 Orgyia pseu
c	73	35.2	2.6	152941	8	AP003044	AP003044 Oryza sat

74	35.2	2.6	162010	2	AP004121	AP004121 Oryza sat
75	35	2.6	388	12	AF316400	AF316400 Synthetic
76	35	2.6	1261	9	AF316401	AF316401 Homo sapi
77	35	2.6	1563	9	AF316402	AF316402 Homo sapi
78	35	2.6	1631	9	BC012134	BC012134 Homo sapi
79	35	2.6	1634	9	BC006413	BC006413 Homo sapi
80	35	2.6	1700	9	AF043250	AF043250 Homo sapi
81	35	2.6	1724	9	BC001779	BC001779 Homo sapi
82	35	2.6	1731	9	BC017224	BC017224 Homo sapi
83	35	2.6	7870	9	HSA275213	AJ275213 Homo sapi
84	35	2.6	11374	1	AE009451	AE009451 Brucella
85	35	2.6	11501	1	AE009276	AE009276 Agrobacte
c 86	35	2.6	12674	1	AE008336	AE008336 Agrobacte
c 87	35	2.6	42655	1	SC7H2	AL109732 Streptomy
c 88	35	2.6	189050	1	AL646066	AL646066 Ralstonia
c 89	35	2.6	213861	9	AL135793	AL135793 Human DNA
90	34.8	2.6	355	11	G72966	G72966 MARC 2567-2
91	34.8	2.6	975	1	AF103943	AF103943 Streptomy
92	34.8	2.6	1325	6	AX168124	AX168124 Sequence
93	34.8	2.6	3105	1	HALHMA	M87833 Halobacteri
c 94	34.8	2.6	10689	1	AE000430	AE000430 Escherich
95	34.8	2.6	14000	6	AX353826	AX353826 Sequence
c 96	34.8	2.6	19088	1	AE001334	AE001334 Chlamydia
c 97	34.8	2.6	225419	1	ECOUW76	U00039 E. coli chr
98	34.6	2.6	1780	8	TAAMY1	X98504 T.aestivum
c 99	34.6	2.6	2968	1	MSU66081	U66081 Mycobacteri
c 100	34.6	2.6	16796	1	AE002269	AE002269 Chlamydia

ALIGNMENTS

```
RESULT   1
A96098
LOCUS       A96098                  1344 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 131 from Patent WO9924578.
ACCESSION   A96098
VERSION     A96098.1  GI:6779941
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1344)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 131 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1344
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      334 a    363 c    362 g    285 t
ORIGIN
```

Query Match 100.0%; Score 1344; DB 6; Length 1344;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 60

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGC 120

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCAC 240

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy     301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGC 360

Qy     361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCT 420

Qy     421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 480

Qy     481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTA 540

Qy     541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTGAGCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCG 600

Qy     601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTG 660

Qy     661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATC 720

Qy     721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAG 780

Qy     781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840

Qy     841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATT 900

Qy     901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTAC 960

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy    1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAA 1080

Qy    1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACT 1140

Qy    1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTC 1200

Qy    1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 1320

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||||||||||||||||||||
Db   1321 GAAACCATTGAGAAGGAAGGCTGA 1344


RESULT   2
A96092
LOCUS       A96092                  1344 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 125 from Patent WO9924578.
ACCESSION   A96092
VERSION     A96092.1  GI:6779938
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1344)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 125 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1344
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      329 a    366 c    365 g    284 t
ORIGIN


  Query Match             90.2%;  Score 1212.8;  DB 6;  Length 1344;
  Best Local Similarity   93.9%;  Pred. No. 3.2e-295;
  Matches 1262;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| | ||
Db      1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 60

Qy     61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
          || |||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||
Db     61 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy    121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
          |||||||||||| |||||||||| || ||||||||||||||||||||||||||||||||||
Db    121 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy    181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
          |||||||||||||||| || |||||||||||||||||||||||||||||||||| ||||||
Db    181 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 240

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy    301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
          ||||| |||||||||| |||||||||||||||||| || ||| ||||| ||||||||||
Db    301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 360

Qy    361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
          |||||||| ||||| || || |||||||||| |||||||||||||||||||||||| |||
Db    361 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 420

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 480

Qy    481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
          |||||||||||||||| |||||||||||||||| || ||||||||||||||||||||||
Db    481 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 540

Qy    541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
          |||||||| |||||||||| |||||||||| ||||| ||||| ||||||||||||||||
Db    541 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 600

Qy    601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
          |||||||||||||||| |||||||||||||||| |||||||||||||||||||| |||||
Db    601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 660

Qy    661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    661 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 720

Qy    721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
          |||||||||| || ||| | || || |||||||| | |||||||||||||| ||||||
Db    721 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 780

Qy    781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          |||||| ||||| | || ||    ||||||||||||||||||||||||||||||||||||
Db    781 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 840

Qy    841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
          |||||||| || || ||| |||| || ||||||||||||||| | ||||||||||||||
Db    841 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 900

Qy    901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
          |||||||||||||||||||| |||||| |||||||||||||| ||||||||||||||||||
Db    901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 960

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy   1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
          |||||||||||||||||||||||||||| || || |||||||||||||| |||||||||
Db   1021 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 1080

Qy   1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
          |||||||||||||  |||||||||||||||||||||||||||||||| ||||| || |||
Db   1081 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 1140

Qy   1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
          || |||||||| ||||| ||||| |||||||| ||| ||||||||||||||||||||||
Db   1141 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 1200

Qy   1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1201 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 1320

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||||||||||||||||||||
Db   1321 GAAACCATTGAGAAGGAAGGCTGA 1344


RESULT   3
AE002412/c
LOCUS       AE002412                12273 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 54 of 206 of
            the complete genome.
ACCESSION   AE002412 AE002098
VERSION     AE002412.1  GI:7225783
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 12273)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
```

```
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 12273)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12273
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            97. .435
                     /gene="NMB0557"
     CDS             97. .435
                     /gene="NMB0557"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40985.1"
                     /db_xref="GI:7225784"
                     /translation="MSDESPIIFTDSCCAKVADLIAEENNPDLKLRVFVNGGGCSGFQ
                     YGFTFDEIKNDDDFEIEKNGLVFLVDPMSYQYLVGAEIDYTESLQGSQFVIRNPNAET
                     TCGCGSSFSV"
     gene            595. .1059
                     /gene="NMB0558"
     CDS             595. .1059
                     /gene="NMB0558"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40986.1"
                     /db_xref="GI:7225785"
                     /translation="MIEMSKDYRNDLYDVYVSYPPQVDRGLIRECLKENLGEEKAEGL
                     IESLDSKPQVLVEEKCTWAKREELHDYFSYLGLDIITRRYMELETVVPPEEGEGEGEG
                     ADGEMPEYLELHGGREDDISAPSQPEPPSRNIKLLVFGLLIAFLGYLLGKIF"
     gene            1094. .2605
                     /gene="NMB0559"
     CDS             1094. .2605
                     /gene="NMB0559"
                     /note="similar to GB:M87049 SP:P27853 PID:148233 GB:U00096
                     PID:2367309 percent identity: 74.73; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ubiquinone biosynthesis protein AarF"
                     /protein_id="AAF40987.1"
                     /db_xref="GI:7225786"
                     /translation="MKWLKRLTVIVGTFYRYRLAGLCASLMGSGWICALLKMMPQSSK
                     LKNEPPAVRLRLALESLGPIFIKFGQVLSTRPDLIPHDYAVELAKLQDKVPPFDARLS
                     REQIEKSLGQSIEKLYAEFETEPIASASIAQVHKARLHSGEQVAVKVLRPNLLPVIEQ
                     DLSLMRFGAGWVERLFADGKRLKPREVVAEFDKYLHDELDLMREAANASQLGRNFQNS
                     DMLIVPKVFYDYCTSDVLTIEWMDGTPVSDIAKLKADGIDLHKLADYGVEIFFTQVFR
                     DGFFHADMHPGNILVAADNRYIALDFGIVGTLTDYDKRYLAINFLAFFNRDYRRVATA
                     HIESGWVPADTRAEELEAAVRAVCEPVFNKPISQISFGLVLMRLFEVSRRFNVEIQPQ
                     LVLLQKTLLNIEGLGRQLDPDLDLWKTAKPFLVKWMNGQVGPKALWRNLKNEAPDWAQ
                     IIPSLPRKISALIDENRQQEMRDAYIHLVKVQQRQSLWLAVIAVVLLLILLLK"
     gene            complement(2962. .3780)
                     /gene="NMB0560"
     CDS             complement(2962. .3780)
                     /gene="NMB0560"
                     /note="similar to GB:L42023 SP:P43886 PID:1005429
                     PID:1220683 PID:1204857 percent identity: 72.97;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="serine acetyltransferase"
                     /protein_id="AAF40988.1"
                     /db_xref="GI:7225787"
                     /translation="MKKDHLNTTGFDLWHTIREETAAAVSAEPMLASFLHQTVLRHES
                     LGSVLAYHLSSKLGSPIMDVRALFEIYQQALGSDTQIGKCVEADLKAIYERDPACDEY
                     SLPLLYFKGFHAIQAHRINHRLYLDGRKTLAYFLQNRMSEVFGVDIHPAARFGYGLML
                     DHATGFVAGETAVLGNNISILHGVTLGGSGKEGGDRHPKIGDGVMIGANASILGNIRI
                     GSNAKIGAGSVVVSDVPPSITVVGVPAKPVARSLKTPSADMDQNIQFTEIDFMI"
     gene            3964. .4542
                     /gene="NMB0561"
     CDS             3964. .4542
                     /gene="NMB0561"
                     /note="similar to GP:3777486 percent identity: 61.03;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="grpE protein"
                     /protein_id="AAF40989.1"
                     /db_xref="GI:7225788"
                     /translation="MSEQTQQQNSEEAVENVEAVETVETVGNADGVQEQAAAEPAYED
                     LQARIAELEAQLKDEQLRALANEQNLRRRHQQEIADTHKFAGQKFAVEMLPVKDYLEM
                     ALLDQSGNFDALKMGVQMTLNELQKAFDATQIKEINPKAGDKLDPNIHQAMQAVASEQ
                     EPNTVVGVMKKGYTLSDRVLRPAMVTVAQKEA"
     gene            complement(4819. .5034)
                     /gene="NMB0562"
     CDS             complement(4819. .5034)
                     /gene="NMB0562"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40990.1"
                     /db_xref="GI:7225789"
                     /translation="MKTLLLTFGIFLTVIIGMAVGYIFSKRSIKGSCGGITALGMKKM
                     CDCDTPCDTLQKKLDEEKQAGGIRVDR"
     gene            complement(5060. .6115)
                     /gene="NMB0563"
     CDS             complement(5060. .6115)
                     /gene="NMB0563"
                     /note="similar to GB:L42023 SP:P44550 PID:1003244
                     PID:1222087 PID:1204428 percent identity: 74.70;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="thiamine biosynthesis lipoprotein ApbE"
                     /protein_id="AAF40991.1"
                     /db_xref="GI:7225790"
                     /translation="MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTT
                     YTVKYLSNNRDKLPSPAEIQKRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRIS
                     SDFAHVTAEAVRLNRLTHGALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGI
                     DKIILKQGKDYASLSKTHPKAYLDLSSIAKGFGVDKVAGELEKYGIQNYLVEIGGELH
                     GKGKNARGEPWRIGIEQPNIVQGGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHI
                     INPNNKRPISHNLASISVVADSAMTADGLSTGLFVLGETEALKLAEREKLAVFLIVRD
                     KGGYRTAMSSEFEKLLR"
     gene            complement(6268. .7485)
                     /gene="NMB0564"
     CDS             complement(6268. .7485)
                     /gene="NMB0564"
                     /note="similar to GB:L42023 PID:1003242 PID:1222086
```

```
                PID:1204427 PID:1573127 percent identity: 92.82;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="Na(+)-translocating NADH-quinone reductase,
                subunit F"
                /protein_id="AAF40992.1"
                /db_xref="GI:7225791"
                /translation="MEIILGIVMFTVIVLVLALMILFAKSKLVSEGDITIKVNGEKEL
                TMPAGGKLLGALANEGIFIPSACGGGGSCGQCRVVVKSGGGDILPTELSHISKREARE
                GCRLSCQVNVKTDMDIEVPEEVFGVKKWECTVISNDNKATFIKELKLAIPEGEEVPFR
                AGGYIQIEAPPHTVAYKDFDIPKEYHEDWDKYNLWQYVSKVDEPILRAYSMASYPEEK
                GIIMLNVRIATPPPRVPDAPPGQMSSYIWSLKPGDKVTISGPFGEFFAKDTDAEMVFI
                GGGAGMAPMRSHIFDQLKRLNSKRKITFWYGARSKREMFYVEDFDQLAAEFPNFTWHV
                ALSDPLPEDNWDGYTGFIHNVVYENHLKNHEAPEDCEFYMCGPPIMNQSVIKMLKDLG
                VEDENILLDDFGG"
     gene       complement(7499. .8092)
                /gene="NMB0565"
     CDS        complement(7499. .8092)
                /gene="NMB0565"
                /note="similar to PID:1573126 percent identity: 91.92;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="Na(+)-translocating NADH-quinone reductase,
                subunit E"
                /protein_id="AAF40993.1"
                /db_xref="GI:7225792"
                /translation="MEHYLSLFIKSVFIENMALSFFLGMCTFLAVSKKVSTAFGLGVA
                VIFVLGLSVPVNQLVYSLLKDGAIAEGVDLTFLKFITFIGVIAALVQILEMFLDKFVP
                ALYNALGIYLPLITVNCAIFGAVSFMAQREYNFGESVVYGFGAGLGWMLAIVALAGIT
                EKMKYSDAPKGLKGLGITFIAAGLMAMAFMSFSGIQL"
     gene       complement(8096. .8722)
                /gene="NMB0566"
     CDS        complement(8096. .8722)
                /gene="NMB0566"
                /note="similar to GB:L42023 SP:P43958 PID:1003237
                PID:1222083 PID:1204425 percent identity: 89.42;

  Query Match             90.2%;  Score 1212.8;  DB 1;  Length 12273;
  Best Local Similarity   93.9%;  Pred. No. 3.8e-295;
  Matches 1262;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| | ||
Db   12069 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 12010

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
           || |||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||
Db   12009 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 11950

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           |||||||||||| |||||||||| || ||||||||||||||||||||||||||||||||||||
Db   11949 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 11890

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
           |||||||||||||||| || |||||||||||||||||||||||||||||||||| ||||||
Db   11889 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 11830

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11829 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 11770

Qy     301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
           |||||| |||||||||| |||||||||||||||| || ||| ||||| ||||||||||||
Db   11769 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 11710

Qy     361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
           |||||||| ||||| || || |||||||||||| |||||||||||||||||||||||| |||
Db   11709 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 11650

Qy     421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   11649 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 11590

Qy     481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
           |||||||||||||||||| |||||||||||||||||| || ||||||||||||||||||||
Db   11589 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 11530

Qy     541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
           ||||||||  |||||||||| |||||||||| ||||| ||||| ||||||||||||||||
Db   11529 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 11470

Qy     601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
           |||||||||||||||| |||||||||||||||| |||||||||||||||||||||| |||
Db   11469 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 11410

Qy     661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   11409 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 11350

Qy     721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaatacccgag 780
           |||||||||| || ||| | || || |||||||| | ||||||||||||||| ||||||
Db   11349 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 11290

Qy     781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
           |||||| ||||| | || ||    ||||||||||||||||||||||||||||||||||||
Db   11289 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 11230

Qy     841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
           |||||||| || || ||| |||| || |||||||||||||| | ||||||||||||||||
Db   11229 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 11170

Qy     901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
           |||||||||||||||||| |||||| |||||||||||| ||||||||||||||||||||
Db   11169 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 11110

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11109 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 11050

Qy    1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
           |||||||||||||||||||||||||||||| || || |||||||||||||| |||||||||
Db   11049 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 10990

Qy    1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
           |||||||||||| ||||||||||||||||||||||||||||||||| ||||| || |||
Db   10989 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 10930

Qy    1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
           || |||||||| |||||| ||||||||| ||| ||||||||||||||||||||||||||
Db   10929 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 10870

Qy    1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10869 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 10810

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10809 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 10750

Qy    1321 gaaaccattgagaaggaaggctga 1344
           |||||||||||||||||||||||||
Db   10749 GAAACCATTGAGAAGGAAGGCTGA 10726


RESULT   4
AX044029/c
LOCUS       AX044029               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 108 from Patent WO0066791.
ACCESSION   AX044029
VERSION     AX044029.1  GI:11342913
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
```

```
          Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 108 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    84410 a  84863 c  94187 g  86520 t
ORIGIN


  Query Match             90.2%;  Score 1212.8;  DB 6;  Length 349980;
  Best Local Similarity   93.9%;  Pred. No. 4.8e-295;
  Matches 1262;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||| | ||
Db 297909 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 297850

Qy     61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
          || ||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||
Db 297849 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 297790

Qy    121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
          |||||||||||| |||||||||| || |||||||||||||||||||||||||||||||||||
Db 297789 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 297730

Qy    181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
          |||||||||||||||| || |||||||||||||||||||||||||||||||||| ||||||
Db 297729 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGGCGCGATTCAC 297670

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297669 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 297610

Qy    301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
          ||||| |||||||||| |||||||||||||||||| || ||| ||||| ||||||||||||
Db 297609 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 297550

Qy    361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
          |||||||| ||||| || || |||||||||| |||||||||||||||||||| ||| |||
Db 297549 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 297490

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 297489 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCT 297430

Qy    481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
          |||||||||||||||| |||||||||||||||| || |||||||||||||||||||||||
Db 297429 GCCGACCCTACGGTCATTATCAAAGAAGCCGCCGAGGATTTCAAACGCGGCCTGTTGGTA 297370

Qy    541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
          ||||||||  |||||||||| |||||||||| ||||| ||||| ||||||||||||||||
Db 297369 TTGAGCCGTTTGACCGAACGCAAAATCCATGTTTGTAAGGCAGCTGGCGCAGACGTGCCG 297310

Qy    601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
          |||||||||||||||| |||||||||||||||| |||||||||||||||||||||| |||
Db 297309 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCTGCCGGTTTG 297250

Qy    661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db 297249 AGTGGCACGCACATTCATTTCATCGAGCCGGTCGGCGCGAATAAAACCGTGTGGACCATC 297190

Qy    721 aattatcaagacgtgattgctatcggacgtttgttctgtaacaggccgtctgaataccgag 780
          |||||||||||| || ||| | || || |||||||| | ||||||||||||||| ||||||
Db 297189 AATTATCAAGATGTAATTACCATTGGCCGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 297130

Qy    781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          |||||| |||| | || ||    ||||||||||||||||||||||||||||||||||||||
Db 297129 CGCGTGATTGCCCTAGGTGGTTCTCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTG 297070

Qy    841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
          |||||||| || || ||| |||| || ||||||||||||||| | ||||||||||||||||
Db 297069 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACACAGACAACCGCGTGATT 297010

Qy    901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
          |||||||||||||||||||| |||||| |||||||||||| ||||||||||||||||||||
Db 297009 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 296950

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296949 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 296890

Qy   1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
          |||||||||||||||||||||||||||| || || |||||||||||||| |||||||||
Db 296889 CCGCAGCCGGACAAATACTCCATCACGCGTACAACCCTCGGCCATTTCCTGAAAAACAAA 296830

Qy   1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
          ||||||||||||  ||||||||||||||||||||||||||||||||| ||||| || |||
Db 296829 CTCTTCAAGTTCAACACAGCCGTCAACGGCGGCGACCGCGCCATGGTGCCGATTGGTACT 296770

Qy   1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
          || |||||||| ||||| ||||||||| ||| ||||||||||||||||||||||||||||
Db 296769 TACGAGCGCGTGATGCCCTTGGATATCCTGCCCACCCTGCTTTTGCGCGATTTAATCGTC 296710

Qy   1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db 296709 GGCGATACCGACAGCGCGCAGGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 296650

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296649 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTG 296590

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||||||||||||||||||||
Db 296589 GAAACCATTGAGAAGGAAGGCTGA 296566


RESULT   5
NMA3Z2491/c
LOCUS       NMA3Z2491             311321 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 3/7.
ACCESSION   AL162754 AL157959
VERSION     AL162754.2  GI:7379424
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 311321)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 311321)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
          sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
          Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT   Notes:
          Details of N. meningitidis sequencing at the Sanger Centre are
          available on the World Wide Web.
          (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .311321
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     RBS             115. .120
     gene            129. .587
                     /gene="NMA0698"
     CDS             129. .587
                     /gene="NMA0698"
                     /note="NMA0698, possible periplasmic protein, len: 152 aa;
                     unknown, contains a probable N-terminal signal sequence
                     and lies within a region of unusually low GC content"
                     /codon_start=1
                     /transl_table=11
                     /product="putative periplasmic protein"
                     /protein_id="CAB83985.1"
                     /db_xref="GI:7379425"
                     /db_xref="SPTREMBL:Q9JQQ2"
                     /translation="MLTKLKILLFLFLFVFVLAINLLFFFFSSDIESFGNYQFEYVYD
                     KGWPANYILVMKDGNEGNFDKIISGLVLEYYKEDDNIYFSYIDGQGFASDSCYYKPEI
                     LYGKIILNKNHIININSMEKNNFLSEDKIMKGTRNWLADPKNKCNIQTLD"
     gene            742. .1077
                     /gene="NMA0699"
     CDS             742. .1077
                     /gene="NMA0699"
                     /note="NMA0699, pseudogene, probable ABC transporter
                     protein, len: 336 bp; similar to C-termini of many ABC
                     transporters e.g. TR:Q46973 (EMBL:U47048), mtfB,
                     Escherichia coli microcin transport protein (707 aa),
                     fasta scores; E(): 2.6e-19, 53.2% identity in 111 aa
                     overlap. Contains PS00211 ABC transporters family
                     signature. NMA0686 may be the remainder of this
                     pseudogene"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="putative ABC transporter protein (pseudogene)"
     misc_feature    868. .912
                     /gene="NMA0699"
                     /note="PS00211 ABC transporters family signature"
     gene            complement(1101. .2327)
                     /gene="NMA0700"
     CDS             complement(1101. .2327)
                     /gene="NMA0700"
                     /note="NMA0700, possible ribonuclease BN, len: 408 aa;
                     similar to SW:RBN_ECOLI (EMBL:L19201), rbn, Escherichia
                     coli ribonuclease BN (EC 3.1.-.-), fasta scores; E():
                     2.3e-28, 34.3% identity in 268 aa overlap. Longer than rbn
                     at the C-terminus. Also similar to TR:O85449
                     (EMBL:AF067083) Vitreoscilla sp. hypothetical protein (376
                     aa), fasta scores; E(): 0, 45.0% identity in 333 aa
                     overlap. Contains a region similar to NMA0677, fasta
                     scores; E(): 4.3e-08, 64.4% identity in 45 aa overlap.
                     Contains hydrophobic, possible membrane-spanning regions,
                     as does rbn"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribonuclease BN"
                     /protein_id="CAB83987.1"
                     /db_xref="GI:7379426"
                     /db_xref="SPTREMBL:Q9JVT9"
                     /translation="MTFLQRLQGLADNKICAFAWFVVRRFDEERVPQAAASMTFTTLL
                     ALVPVLTVMVAVASIFPVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTA
                     IGSVMLVVTSLMLIRTIDNTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFM
                     VGSVQDAALASGAPQWSGALRTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATA
                     FCLETARSLFTWYMGNFDGYRSIYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQ
                     GEAFRRGFDSRGRFDDVLKILLLLDAAQKEGKALPVQEFRRHINMGYDELGELLEKLA
                     RHGYIYSGRQGWVLKTGADSIELNELFKLFVYRPLPVERDHVNQAVDAVMMPCLQTLN
                     MTLAEFDAQAKKQQQS"
     misc_feature    1376. .1385
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(2417. .2426)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2444. .2453
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            2486. .2743
                     /gene="NMA0701"
     CDS             2486. .2743
                     /gene="NMA0701"
                     /note="NMA0701, possible pseudogene, len: 258 bp; shows
                     weak similarity to part of SW:WRBA_ECOLI (EMBL:M99166),
                     wrbA, Escherichia coli Trp repressor binding protein (197
                     aa), fasta scores; E(): 0.064, 29.3% identity in 82 aa
                     overlap. Also similar to part of TR:O85450 (EMBL:AF067083)
                     Vitreoscilla sp. Trp repressor binding protein (fragment)
                     (124 aa), fasta scores; E(): 6.6e-14, 56.6% identity in 83
                     aa overlap"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="hypothetical protein NMA0701 (pseudogene)"
     misc_feature    complement(2755. .2764)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    2778. .2787
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             2861. .2864
     gene            2871. .3530
                     /gene="NMA0702"
     CDS             2871. .3530
                     /gene="NMA0702"
                     /note="NMA0702, len: 219 aa; similar to TR:Q52925
                     (EMBL:Z50189), exsB, Rhizobium meliloti putative regulator
                     of succinoglycan biosynthesis (not a transcriptional
                     regulator) (234 aa), fasta scores; E(): 3.1e-15, 32.7%
                     identity in 205 aa overlap. Also similar to many
                     hbacterial ypothetical proteins e.g. SW:YBAX_HAEIN
                     (EMBL:U32798), HI1191, Haemophilus influenzae hypothetical
                     protein (196 aa), fasta scores; E(): 0, 79.2% identity in
                     173 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0702"
                     /protein_id="CAB83989.1"
                     /db_xref="GI:7379427"
                     /db_xref="SPTREMBL:Q9JVT8"
                     /translation="MSNQKALVIFSGGQDSTTCLIQAIQTYGRENVQAITFQYGQRHA
                     VELERARWIAQDLGVKQTVLDLSLMRQITHNALMDDTAAIETAENGVPNTFVDGRNAL
                     FLLYAAIYAKGQGIRHIIAGVCETDFSGYPDCRDVFVKSMNVTLNLAMDYDFQIHTPL
                     MYLTKAQTWALADEMGALDYIREQTHTCYNGIVGGCRECPSCILRERGLAEYLESKKA
                     V"
     misc_feature    3522. .3531
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             3547. .3550
     gene            3558. .4076
                     /gene="NMA0703"
     CDS             3558. .4076
                     /gene="NMA0703"
                     /note="NMA0703, len: 172 aa; unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0703"
```

```
                     /protein_id="CAB83990.1"
                     /db_xref="GI:7379428"
                     /db_xref="SPTREMBL:Q9JVT7"
                     /translation="MPKLHMFYLGGNAGRSNIEVHDIQFAVCDDYREAVPALKAAWFG
                     DADKIHIDGWQVVEWADGYDIAVSETPKTKMPSENAPRLYFANVGGYRAGQLAEAHAF
                     GLFAAATPAEAKQKALQTLLTDSYVQQHKDNLKDVDNLLALDRIGNFHIRLTPNPHGK
                     PAEIGFQGYLPI"
     misc_feature    3782. .3791
                     /gene="NMA0703"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            4083. .4505
                     /gene="NMA0704"
     misc_feature    4083. .4502
                     /gene="NMA0704"
                     /note="Pfam match to entry PF01242 PTPS, 6-pyruvoyl
                     tetrahydropterin synthase, score 10.80, E-value 1.7e-06"
     CDS             4083. .4505
                     /gene="NMA0704"
                     /note="NMA0704, len: 140 aa; similar to many hypothetical
                     proteins e.g. SW:YB90_HAEIN (EMBL:U32798), HI1190,
                     Haemophilus influenzae hypothetical protein (141 aa),
                     fasta scores; E(): 5.1e-32, 58.0% identity in 138 aa
                     overlap. Shows very weak similarity to eukaryotic
                     6-pyruvoyl-tetrahydropterin synthases e.g. SW:PTPS_RAT
                     (EMBL:M77850), pts, Rattus norvegicus 6-pyruvoyl
                     tetrahydrobiopterin synthase precursor (EC 4.6.1.10) (144
                     aa), fasta scores; E(): 0.0062, 27.3% identity in 154 aa
                     overlap. Contains Pfam match to entry PF01242 PTPS,
                     6-pyruvoyl tetrahydropterin synthase"

  Query Match             87.7%;  Score 1179.2;  DB 1;  Length 311321;
  Best Local Similarity   92.3%;  Pred. No. 1.4e-286;
  Matches 1241;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  55476 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 55417

Qy     61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
          |||||||| || | ||||||||||||||||||||||||||||||||| ||| ||||||
Db  55416 TATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 55357

Qy    121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
          |||||||||||| |||||||||| || ||||||||||||||||||||||||||||||||
Db  55356 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 55297

Qy    181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
          |||||||||||||||| || ||||| ||||||| |||||||||||||||| || ||
Db  55296 AAAAAGAATCCGGGCGTGGTGTTTACCGCGCCGGTTTCAGGCAAAATCGCCGCCATCCAT 55237

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          || |||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db  55236 CGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 55177

Qy    301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
          ||||||||||||||||  || |||||| ||||||| || ||| ||||| ||||||||||
Db  55176 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 55117

Qy    361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
          || ||||| ||||| || || |||||||||||| || ||||||||||||||||||||||
Db  55116 AATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTACCCGTCCGTTCAGCAAAATCCCT 55057

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55056 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCG 54997

Qy    481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
          || ||||||  |||  | |||||||||||||||| || |||| ||| ||  || ||||||
Db  54996 GCAGACCCTGTGGTTGTGATCAAAGAAGCCGCCGAGGATTTCAGACGAGGTTTGCTGGTA 54937

Qy    541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
          ||||||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||

Db  54936 TTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 54877

Qy    601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
          |||||||||||||||| |||||||||||||||| |||||||||||||| ||||| |||
Db  54876 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTG 54817

Qy    661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
          |||||||||||||||||||||||| ||||| ||||| || || |||||||| |||||||||
Db  54816 AGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATC 54757

Qy    721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
          |||||||||||  || ||||| |||||||||||||| | |||||||||||||| ||||||
Db  54756 AATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 54697

Qy    781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          |||||| |||| ||||| ||    |||||||||||||| ||||||||||||||||||||
Db  54696 CGCGTGATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTG 54637

Qy    841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
          |||||||| || || ||| |||| || ||||||||||||||||| ||||||||||||||
Db  54636 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 54577

Qy    901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
          |||||||||||||||||||| |||||| |||||||||||| ||||||||||||||||||
Db  54576 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 54517

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54516 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 54457

Qy   1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
          |||||||||||||||||||||||||||||||| || || |||||||||||||| |||||||||
Db  54456 CCGCAGCCGGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAA 54397

Qy   1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
          |||||||||||||||||||||||||||||| |||||||||||||||| ||||| || |||
Db  54396 CTCTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACT 54337

Qy   1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
          || |||||||||||||| | |||||||||||||| |||||||||||||||||||||||||
Db  54336 TACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 54277

Qy   1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
          |||||||||||||||||||| || ||||||||||||||||||||||||||||||||||||
Db  54276 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 54217

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||| || ||||||
Db  54216 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGTAAGGTGCTG 54157

Qy   1321 gaaaccattgagaaggaaggctga 1344
          |||||||||||||||||||||||||
Db  54156 GAAACCATTGAGAAGGAAGGCTGA 54133


RESULT   6
A96094
LOCUS       A96094                  1344 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 127 from Patent WO9924578.
ACCESSION   A96094
VERSION     A96094.1  GI:6779939
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1344)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 127 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .1344
```

```
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      323 a    347 c    368 g    289 t     17 others
ORIGIN


  Query Match             86.7%;  Score 1165.4;  DB 6;  Length 1344;
  Best Local Similarity   91.2%;  Pred. No. 2.9e-283;
  Matches 1226;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGTCATT 60

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
           |||||||| || | ||||||||||||||||||||||||||||||||||| ||| ||||||
Db      61 TATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           |||| ||||||| |||||||||| || ||||||||||||||||||||||||||||||||||
Db     121 CCCTNGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
           |||||| |||||||||| || ||||| ||||| | ||||||||||||||||||| || ||
Db     181 AAAAAGNATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCAT 240

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           || |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     241 CGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300

Qy     301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
           ||||||||||||||  || |||||| ||||||| || ||| ||||  || |  |  ||
Db     301 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGAANTNNGNNGC 360

Qy     361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
           || ||||| ||||| || || |||||||||| || | ||||||||||||||||||||||
Db     361 AATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCT 420

Qy     421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGCG 480

Qy     481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
           || |||||| ||| | |||||||||||| ||| || |||| ||| |  || |||||
Db     481 GCAGACCCTGTGGTTGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTA 540

Qy     541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
           ||||||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db     541 TTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 600

Qy     601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
           |||||||||||||||| |||||||||||||||| |||||||||||||| ||||| |||
Db     601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTG 660

Qy     661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
           ||||||||||||||||||||||| ||||| ||||| || || |||||||| |||||||||
Db     661 AGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATC 720

Qy     721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
           |||||||||||  || ||||| |||||||||||||| | ||||||||||||||| ||||||
Db     721 AATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAG 780

Qy     781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
           |||||| |||| |||||  ||    |||||||||||||| ||||||||||||||||||||
Db     781 CGCGTGATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTG 840

Qy     841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
           |||||||| || || ||| |||| || ||||||||||||||||| ||||||||||||||
Db     841 GGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900

Qy     901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
           |||||||||||||||||||| |||||| |||||||||||||| ||||||||||||||||||
Db     901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 960

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020

Qy    1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
           |||||||||||||||||||||||||||| || || |||||||||||||| |||||||||
Db    1021 CCGCAGCCGGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAA 1080

Qy    1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
           |||||||||||||||||||||||||||||| |||||||||||||||| ||||| || |||
Db    1081 CTCTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACT 1140

Qy    1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
           || |||||||||||||| | |||||||||||||| |||||||||||||||||||||||||
Db    1141 TACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 1200

Qy    1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
           |||||||||||||||||||| || ||||||||||||||||||||||||||||||||||||
Db    1201 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT 1260

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db    1261 TTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCCCGCTGTTGCGTAAGGTGCTG 1320

Qy    1321 gaaaccattgagaaggaaggctga 1344
           |||||| ||||||||||||||||||
Db    1321 GAAACCNTTGAGAAGGAAGGCTGA 1344


RESULT   7
AE006171
LOCUS       AE006171               10563 bp    DNA     linear   BCT 08-MAR-2001
DEFINITION  Pasteurella multocida PM70 section 138 of 204 of the complete
            genome.
ACCESSION   AE006171 AE004439
VERSION     AE006171.1  GI:12721691
KEYWORDS    .
SOURCE      Pasteurella multocida.
  ORGANISM  Pasteurella multocida
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Pasteurella.
REFERENCE   1  (bases 1 to 10563)
  AUTHORS   May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
            Kapur,V.
  TITLE     Complete genomic sequence of Pasteurella multocida,Pm70
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
  MEDLINE   21145866
REFERENCE   2  (bases 1 to 10563)
  AUTHORS   Zhang,Q. and Kapur,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
            University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
            55108, USA
FEATURES             Location/Qualifiers
     source          1. .10563
                     /organism="Pasteurella multocida"
                     /strain="PM70"
                     /db_xref="taxon:747"
     gene            94. .1605
                     /gene="PM1326"
     CDS             94. .1605
                     /gene="PM1326"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK03410.1"
                     /db_xref="GI:12721692"
                     /translation="MWIDVMTQRDCILKLKNIGKSFFDVPVLKNVDLDIYAGEVHCLV
                     GENGAGKSTLCKIIAGIYPPEQGEMFYEGKSYFPKTVKEAQSLGIGFIHQELMLVPQL
                     SVLENIFLGNEKTVFCGKMDWQTMRHKTQQIISELELDIDPDAKVADLSIAQQQMVEI
                     AKAVFSEYKVIIFDEPTSSISRKNTQTLFKIIHQLKEKNVAMIYISHRLDEFQYIADQ
                     VTVLRDGIRTGTMAYRDTSPDEIVKLMVGREINFAEYHRHYLPTREVLRVESLTNKHI
```

```
                KSSSFSLYEGEILGFAGLVGAGRTELLRAIFGADEASGAIYLNAQRVEIRSPEDAVKQ
                KIGFITEDRKQQGLILGLSIRENITLPILKRFWNGLFLNKKQERLVAEENRQKLRIVS
                HSQDQISQTLSGGNQQKVIIARWLESGVKILFFDEPTRGIDVGAKSEIYDLMRQFTDN
                GGAIVMVSSDLPELITMSDRVIVMRQGGIVREVKDKAEMTEENLMRLMIGVDE"
gene            1598. .2641
                /gene="rbsC_3"
                /note="PM1327"
CDS             1598. .2641
                /gene="rbsC_3"
                /codon_start=1
                /transl_table=11
                /product="RbsC"
                /protein_id="AAK03411.1"
                /db_xref="GI:12721693"
                /translation="MNRYCMTNLRKIFSKLGIGIILLLMIIGMSLSSDVFLSTNNIIN
                ILLQVSIICVISVGMTYVILTGGIDLSVGSIVALSAVCLGVFTHWGMDWLGENPSSFS
                VLMIVISAIIATIFVGILCGYVNGLVIVYGKVTPFITTLGMMGIARGLALTISDGKTI
                YNFPDTLRFLGNGRIALTETFALPVPVIIALLVVLVSFYVLTQTMFGRQIYALGGNRE
                AVRLSGINIEKLEIKAYVINGALAAIGAIILVGRLNAAQPIAGNGYELDAIAATVIGG
                TSLMGGVGSVVSTSIGALIMGVLQNGLTLLNVTSYLQRLIIGLVIILAVFLDQLRRGE
                VSTRRLKRLFFRE"
gene            3187. .4527
                /gene="PM1328"
CDS             3187. .4527
                /gene="PM1328"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03412.1"
                /db_xref="GI:12721694"
                /translation="MITIKKGLNLPISGSPEQVIRDGNAITEVALLGEEYVGMRPSMK
                VREGDVVKKGQVLFEDKKNPGVVFTAPASGTVTAIHRGAKRVLQSVVIKIEGNEQITF
                EKYTTEQLNQLTSEQVRQNLQASGLWTALRTRPFSKVPAVDATPVSIFVNAMDTNPLC
                ADPQVIVQQSAQAFEAGLTVLSRLHEGKVYLCKAANASIPSPSIANLDVKEFAGPHPA
                GLSGTHIHFIDPVSATKFVWYINYQDVIAVGKLFTTGELDVSRVVSLAGPQVKNPRLV
                RTVLGANLSQLTANEVKDGENRVISGSVLSGAKAAGPVDYLGRYALQVSVLEEGREKE
                FLGWIMPGANKYSLSRTVLGHFSKKLFNFTTALNGGERAMVPIGAYERVMPLDIIPTL
                LLRDLAAGDTDSAQALGCLELDEEDLALCTFVCPGKNEYGPLLRQALDKIEKEG"
gene            4530. .5762
                /gene="nqrB"
                /note="PM1329"
CDS             4530. .5762
                /gene="nqrB"
                /codon_start=1
                /transl_table=11
                /product="NqrB"
                /protein_id="AAK03413.1"
                /db_xref="GI:12721695"
                /translation="MGLKHFIEKIEPAFLPGGKYEKWYALFEATATFLYTPGTVTHKA
                SHVRDALDSKRMMVLVWLSLFPAMFYGMYNVGAQAILATDALGTLQQAISGHWQYALS
                HALGADLTLSAGWASKMLLGATYFLPIYLTIFLVGGFWEVLFAMVRKHEINEGFFVTS
                ILLALIVPPTLPLWQAALATTFGVVVAKEIFGGVGKNFMNPALAARAFLFFAYPAQIS
                GDTVWVAADGFSGATALSQWAVGGEAGLKHVVTGQPITWMDAFIGNIPGSIGEVSTLA
                LIIGAAIIVFARIASWRIIAGVMIGMIATSALFNLIGSSTNPLFAMPWYWHFVLGGFA
                LGMFFMATDPVSAAFTNKGKWWYGILIGAMAVLIRVVNPAYPEGMMLAILFANLFAPV
                FDYLVVQANIKRRRARNV"
gene            5755. .6537
                /gene="PM1330"
CDS             5755. .6537
                /gene="PM1330"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03414.1"
                /db_xref="GI:12721696"
                /translation="MFKNKDSVGGTLLVIILLSLACSIIVAGSAVLLKPTQIEQKELD
                KQKNILSVAGLLQPTTKNSEIKTIYANNIEARLVDLNTGDFAPAQPGFDAAKAVKNPA
                ESTALSAEDDVAGIRVRANFAEVYLVKNDAGQVTQVVLPFYGKGLWSMMYGFMSVQPD
                GNTVNGITYYDQGETPGLGGEIENPKWQAQFPGKKLYTADNQVGLYVGKGASANAEHG
                IDAISGSTLTSNGVNNSFKFWLGQKGFGPFLAKLKAGVLNNG"
gene            6530. .7156
                /gene="PM1331"
CDS             6530. .7156
                /gene="PM1331"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03415.1"
                /db_xref="GI:12721697"
                /translation="MADTKKLKGLLLSPVMDNNPIALQILGICSALAVTTKLETAIVM
                TFAVIFVTAFSNLFISLIRNYIPNSVRIIVQMAIIASLVILVDQILRAYAYGLSKQLS
                VFVGLIITNCIVMGRAEAFAMKSQPLESFVDGIGNGLGYGAILVSVAFIRELIGSGKL
                FGMTVFQTIQDGGWYQTNGLFLLAPSAFFIIGFIIWGIRTLKPNQVEK"
gene            7160. .7756
                /gene="nqrE"
                /note="PM1332"
CDS             7160. .7756
                /gene="nqrE"
                /codon_start=1
                /transl_table=11
                /product="NqrE"
                /protein_id="AAK03416.1"
                /db_xref="GI:12721698"
                /translation="MEHYISLFVKSVFIENMALSFFLGMCTFLAVSKKVSTAFGLGIA
                VIVVLGIAVPVNQLVYSFILKDSALVQGIDLSFLNFITFIGVIAALVQILEMVLDKYF
                PALYNALGIFLPLITVNCAIFGGVSFMVQRDYTFVESVVYGIGAGTGWMLAIVALAGI
                TEKMKYADVPAGLRGLGITFITVGLMALGFMSFSGIQL"
gene            7772. .8995
                /gene="nqr6"
                /note="PM1333"
CDS             7772. .8995
                /gene="nqr6"
                /codon_start=1
                /transl_table=11
                /product="Nqr6"
                /protein_id="AAK03417.1"
                /db_xref="GI:12721699"
                /translation="MEITLGIAMFTVIVLALAVIILFAKSKLVNSGDITIEINDDPSK
                AIHLPAGGKLLGALASQGIFVSSACGGGGSCGQCIVKVTEGGGDILPTELSHISKREA
                KEGYRLSCQVNVKNSMKVELPEEVFGVKKWECTVISNDNKATFIKELKLAIPEGEEVP
                FRAGGYIQIEAEPHTVAYKDFDIPEEYHEDWDKYNLWRYVSKVDEHIIRAYSMASYPE
                EKGIIMLNVRIATPPPNNPDAPPGQMSSYIWSLKAGDKVTISGPFGEFFAKETDAEMV
                FVGGGAGMAPMRSHIFDQLKRLKSKRKMSFWYGARSKREMFYVEDFDTLQAENDNFVW
                HVALSDPQPGDNWDGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPIMNASVIKMLKD
                LGVEDENILLDDFGG"
gene            9164. .10204
                /gene="apbE"
                /note="PM1334"
CDS             9164. .10204
                /gene="apbE"
                /codon_start=1
                /transl_table=11
                /product="ApbE"
                /protein_id="AAK03418.1"
                /db_xref="GI:12721700"
                /translation="MKHILNSIIIGFCSLCLLACHKSPEVVTITGKTMGTTYSVKYID
                DGQIKIKPIEMHGQIEEILKEVNNKMSTYLSTSELSQFNQNTQIDTPVEISADLAFVL
                KEAIRLHGVTEGALDVTVGPIVNLWGFGPEKRQPNETLEEQVEKRRAWVGIEKMKLTE
                QGDKFYLAKTVPQLYIDLSSIAKGFGVDKVAAHLATKGIANYLVEIGGEINAKGKNIE
                GKDWQIAIEKPNFDGSRAVSQIIGLKDFSMATSGNYRNYFEENGRRFSHEIDPKTGYP
                IEHRLASITVLADSTMTADGLSTGLYVLGEEKALEVAEKYNLLVYLISKTEKGFEAKM
                SSAFEKLLNEQH"
gene            10214. .10468
                /gene="PM1335"
CDS             10214. .10468
                /gene="PM1335"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAK03419.1"
                /db_xref="GI:12721701"
                /translation="MQLFLITFGIFLLIIAGMAIGYIIQRKTIAGSCGGISALGLKKV
                CDCEEPCDNLKAKLEAGDEEAQQEYNAKFAKNEPHFYEVK"
BASE COUNT      3052 a   1783 c   2451 g   3277 t
ORIGIN
```

```
 Query Match             42.7%;  Score 574.4;  DB 1;  Length 10563;
 Best Local Similarity   64.7%;  Pred. No. 4.6e-134;
 Matches  870;  Conservative    0;  Mismatches  471;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
Db    3187 ATGATTACGATTAAAAAAGGCTTGAATCTTCCTATTTCGGGAAGCCCAGAGCAGGTAATC 3246

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
Db    3247 CGTGATGGCAATGCTATCACTGAAGTTGCTTTGCTTGGTGAAGAGTATGTGGGAATGCGC 3306

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
Db    3307 CCTTCAATGAAGGTGCGTGAAGGTGATGTTGTGAAAAAAGGCCAAGTGCTTTTTGAAGAT 3366

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
Db    3367 AAAAAGAACCCTGGTGTGGTTTTCACAGCCCCTGCAAGTGGTACTGTAACTGCGATTCAT 3426

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
Db    3427 CGTGGTGCTAAGCGAGTTTTACAATCTGTAGTAATTAAAATTGAAGGTAATGAACAAATT 3486

Qy     301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
Db    3487 ACTTTTGAAAAATATACAACAGAACAATTAAACCAACTTACTTCAGAGCAAGTTCGTCAA 3546

Qy     361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
Db    3547 AATTTACAAGCTTCTGGATTATGGACCGCACTTCGTACGCGTCCGTTCAGTAAAGTGCCT 3606

Qy     421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
Db    3607 GCCGTCGATGCCACACCCGTATCGATTTTCGTGAATGCGATGGATACCAATCCATTATGT 3666

Qy     481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
Db    3667 GCCGATCCACAAGTGATCGTGCAACAATCTGCGCAAGCTTTTGAAGCGGGTCTGACTGTA 3726

Qy     541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
Db    3727 TTAAGTCGTTTACATGAGGGGAAAGTGTACCTGTGTAAAGCGGCGAATGCGAGCATTCCA 3786

Qy     601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
Db    3787 AGCCCAAGCATTGCCAATCTTGACGTGAAAGAATTTGCAGGTCCACATCCAGCTGGGTTA 3846

Qy     661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
Db    3847 AGTGGTACGCATATTCACTTTATTGATCCTGTGAGTGCAACCAAATTTGTTTGGTATATC 3906

Qy     721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
Db    3907 AATTACCAAGATGTGATAGCGGTGGGTAAATTATTTACCACAGGTGAATTAGATGTTTCA 3966

Qy     781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
Db    3967 CGCGTAGTGTCCTTAGCTGGTCCACAAGTGAAAAATCCAAGATTAGTACGTACCGTATTA 4026

Qy     841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
Db    4027 GGTGCAAACCTTTCCCAATTAACAGCGAATGAAGTGAAAGACGGTGAAAATCGCGTGATT 4086

Qy     901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
Db    4087 TCAGGTTCAGTATTAAGTGGTGCAAAAGCTGCAGGTCCTGTTGATTATTTAGGTCGTTAC 4146

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db    4147 GCATTGCAAGTTTCAGTACTTGAAGAAGGTCGTGAGAAAGAATTTCTAGGTTGGATCATG 4206

Qy    1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
Db    4207 CCGGGTGCGAATAAATACTCACTTTCACGCACGGTGTTAGGGCATTTCTCTAAA---AAA 4263

Qy    1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
Db    4264 TTATTTAATTTTACCACCGCACTTAATGGCGGTGAGCGTGCGATGGTACCTATCGGTGCT 4323

Qy    1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
Db    4324 TATGAGCGTGTGATGCCATTGGATATTATTCCAACATTATTACTGCGTGATTTGGCAGCA 4383

Qy    1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
Db    4384 GGCGATACCGATTCTGCACAGGCGTTAGGGTGTTTAGAGCTAGATGAAGAAGATTTAGCC 4443

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
Db    4444 CTTTGTACCTTTGTTTGCCCTGGTAAAAATGAATATGGTCCGCTCTTACGTCAAGCTTTA 4503

Qy    1321 gaaaccattgagaaggaaggctga 1344
Db    4504 GATAAGATCGAGAAGGAAGGTTAA 4527
```

```
RESULT   8
APU24492
LOCUS       APU24492                1558 bp    DNA     linear   BCT 23-FEB-1996
DEFINITION  Actinobacillus pleuropneumoniae 48 kDa outer membrane protein
            (aopA) gene, complete cds.
ACCESSION   U24492
VERSION     U24492.1  GI:1185394
KEYWORDS    .
SOURCE      Actinobacillus pleuropneumoniae.
  ORGANISM  Actinobacillus pleuropneumoniae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Actinobacillus.
REFERENCE   1  (bases 1 to 1558)
  AUTHORS   Cruz,W.T., Nedialkov,Y.A., Thacker,B.J. and Mulks,M.H.
  TITLE     Molecular characterization of a common 48-kilodalton outer membrane
            protein of Actinobacillus pleuropneumoniae
  JOURNAL   Infect. Immun. 64 (1), 83-90 (1996)
  MEDLINE   96110919
REFERENCE   2  (bases 1 to 1558)
  AUTHORS   Mulks,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-1995) Martha H. Mulks, Microbiology, Michigan
            State University, 367 Giltner Hall, East Lansing, MI 48824, USA
FEATURES             Location/Qualifiers
     source          1. .1558
                     /organism="Actinobacillus pleuropneumoniae"
                     /strain="ISU158"
                     /serotype="5"
                     /db_xref="taxon:715"
     gene            95. .1444
                     /gene="aopA"
     CDS             95. .1444
                     /gene="aopA"
                     /note="similar to ngrA gene of Vibrio alginolyticus
                     encoding the alpha subunit of a sodium-translocating
                     NADH-uiquinone oxidoreductase, Genbank Accession Number
                     Z37111"
                     /codon_start=1
                     /transl_table=11
                     /product="48 kDa outer membrane protein"
                     /protein_id="AAC43631.1"
                     /db_xref="GI:1185395"
                     /translation="MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMK
                     VREGDVVKKGQVLFEDKKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITF
                     TRYEAAQLASLSAEQVKQNLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLA
                     ADPEVVLKEYETDFKDGLTVLTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGV
                     HPAGLVGTHIHFVDPVGATKQVWHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNP
                     RLVRTRLGANLSQLTANELNAGENRVISGSVLSGATAAGPVDYLGRYALQVSVLAEGR
```

```
                 EKELFGWIMPGSDKFSITRTVLGHFGKKLFNFTTAVHGGERAMVPIGAYERVMPLDII
                 PTLLLRDLAAGDTDSAQNLGCLELDEEDLALCTYVCPGKNNYGPMLRAALEKIEKEG"
     terminator  1462. .1499
BASE COUNT      432 a    299 c    375 g    452 t
ORIGIN
```

```
  Query Match             42.1%;  Score 566.2;  DB 1;  Length 1558;
  Best Local Similarity   65.3%;  Pred. No. 4.7e-132;
  Matches  883;  Conservative    0;  Mismatches  458;  Indels   12;  Gaps    3;

Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| |||||| |||||  |  || | || ||||| || | |||||  ||||| ||
Db      95 ATGATTACAATCAAGAAAGGCTTGGATTTACCTATCGCAGGTACACCGGCACAAGTAATC 154

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
           || |||||    || |||  ||||| ||| ||||||||||||||||||| || |||||
Db     155 CATAACGGCAATACCGTTAATGAAGTTGCGATGCTTGGCGAAGAATATGTGGGTATGCGT 214

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           || || |||||| |    ||||| || |  || |||||||| || || || ||||||||
Db     215 CCTTCAATGAAAGTTCGTGAAGGCGATGTAGTGAAAAAAGGTCAGGTTCTTTTTGAAGAT 274

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
           |||||||||||||||| |||||||||| || ||    || |   | |  |||||  ||
Db     275 AAAAAGAATCCGGGCGTTGTATTTACTGCTCCTGCAAGCGGTACCGTGGTTACTATTAAC 334

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           || || |||||||| || |||||||| |||||||||   |||||  ||  |  ||  ||||
Db     335 CGCGGTGAAAAGCGTGTTCTTCAGTCGGTCGTGATTAAAGTTGAGGGTGATGAGCAAATT 394

Qy     301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
              ||    |||||  |  | |||      | ||     |      |||  |||||   |
Db     395 ACCTTTACTCGCTATGAAGCTGCGCAATTAGCGTCGCTATCAGCCGAGCAAGTGAAACAA 454

Qy     361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
           || || ||  |||| || ||||||||||||  | || || ||||||||||| ||  | ||
Db     455 AATCTTATCGAATCGGGTTTATGGACTGCATTCCGTACTCGTCCGTTCAGTAAGGTTCCC 514

Qy     421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           ||  | || ||    ||||   | ||||| |||||||| ||||| |||||||||  | ||
Db     515 GCATTGGACGCGATTCCGTCATCAATCTTTGTCAATGCAATGGATACCAATCCGTTAGCG 574

Qy     481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
           || || ||    ||  |  | |||||   ||     || || |||  |||  |    ||
Db     575 GCGGATCCGGAAGTGGTTTTAAAAGAGTACGAAACCGATTTTAAAGACGGTTTAACTGTT 634

Qy     541 ttgagccgcctgaccgaacgtaaaa------tccatgtgtgtaaagcagcaggcgcagac 594
           || | |||  |     |  || |||      |  |  | |||||  || | |    |
Db     635 TTAACCCGTTTATTTAACGGTCAAAAACCGGTTTACTTATGTAAAGACGCGGACAGCAAT 694

Qy     595 gtgccgtctgaaaatgc---tgccaatatcgaaacacatgaatttggcggcccgcatcct 651
             | ||||        ||   ||    ||||  ||    |   ||| ||||    ||||||
Db     695 ATCCCGTTAAGTCCGGCGATTGAAGGTATCACAATTAAGTCGTTTAGCGGTGTTCATCCT 754

Qy     652 gccggcttgagtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtg 711
           |||||  ||    || ||||| || || ||  | || || || || |||| |||   ||
Db     755 GCCGGTTTAGTCGGTACGCATATCCACTTTGTTGATCCCGTAGGGGCGACCAAACAAGTT 814

Qy     712 tggaccatcaattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctg 771
           |||  | | ||||||||||| ||||||||||||||    || ||    ||||||   ||
Db     815 TGGCACTTAAATTATCAAGATGTGATTGCTATCGGTAAATTATTTACTACAGGCGAACTC 874

Qy     772 aataccgagcgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgt 831
             ||||||  ||  |  || |  | | ||| | |||||| || || || ||  |  | |||
Db     875 TTTACCGACCGTATTATTTCGCTTGCCGGTCCGCAAGTGAAAAATCCTCGTTTAGTACGT 934

Qy     832 accgttttgggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaac 891
           ||   | | ||||||||  | || ||| | |||||    || ||   || ||  || |||
Db     935 ACGCGTCTTGGTGCGAATCTTTCCCAATTAACCGCAAATGAGTTAAATGCCGGTGAAAAC 994

Qy     892 cgcgtgatttccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttg 951
           || |||||||| |||||||||||||| |||||||| |||    ||  ||  |||||| ||
Db     995 CGTGTGATTTCAGGTTCGGTATTGAGCGGTGCGACTGCGGCGGGTCCGGTTGATTACTTA 1054

Qy     952 ggacgctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggc 1011
           || || |||      ||  | |||||  | |  |||||  ||    |||||| | |||||
Db    1055 GGCCGTTACGCATTACAAGTATCCGTACTGGCGGAAGGTCGTGAGAAAGAGTTATTCGGT 1114

Qy    1012 tgggttgcgccgcagccggacaaatactccatcacgcgcaccactctcggccatttccta 1071
           ||| ||  ||||    |||| |||| |||||| || || |||    | ||||| |||
Db    1115 TGGATTATGCCGGGTTCGGATAAATTCTCCATTACCCGTACCGTGTTAGGCCACTTCGGT 1174

Qy    1072 aaaaacaaactcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccg 1131
           |||   ||| | || || ||||| || || ||  |||||||  || || || ||||| ||
Db    1175 AAA---AAATTATTTAACTTCACTACGGCAGTTCACGGCGGTGAACGTGCAATGGTGCCA 1231

Qy    1132 atcggcacttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgat 1191
           |||||  | |||||||| |||||||||| || ||  | ||||| ||  |  | || |||
Db    1232 ATCGGTGCGTATGAGCGTGTAATGCCGTTAGATATTATTCCTACGTTATTACTTCGTGAT 1291

Qy    1192 ttaatcgtcggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaa 1251
           |||   | ||||||||||||    || ||    || ||||| || ||  |||| ||||||
Db    1292 TTAGCTGCCGGCGATACCGATTCTGCACAAAACTTAGGTTGTTTAGAGCTGGATGAAGAA 1351

Qy    1252 gacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgc 1311
           ||  | || || || |  |  || |||||||| ||| |  | ||||| ||| |||| ||
Db    1352 GATTTGGCATTATGTACTTATGTTTGCCCGGGTAAAAATAACTACGGTCCGATGTTACGT 1411

Qy    1312 aaagtgctggaaaccattgagaaggaaggctga 1344
             | | || |  || ||||||||||| | |
Db    1412 GCTGCGTTAGAGAAGATCGAGAAGGAAGGTTAA 1444
```

```
RESULT   9
U32702
LOCUS       U32702                 13299 bp    DNA     linear   BCT 29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 17 of 163 of the complete genome.
ACCESSION   U32702 L42023
VERSION     U32702.1  GI:1573118
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
  ORGANISM  Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 13299)
  AUTHORS   Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
REFERENCE   2  (bases 1 to 13299)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
REFERENCE   3  (bases 1 to 13299)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 13299)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
```

```
  REMARK     The H. influenzae sequence has been updated by R. Fleischmann. New
             database matches have been assigned, product names have been
             improved, and a number of frame shifts have been corrected. We
             gratefully acknowledge the work of Tatusov et. al. We have
             incorporated their annotation into the /notes fields of the
             corresponding H. influenzae genes
REFERENCE    5  (bases 1 to 13299)
  AUTHORS    White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
             Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
             Medical Center Dr, Rockville, MD 20850, USA
  REMARK     The whole genome was shifted by 588 nucleotides for a new start
COMMENT      On Sep 30, 1996 this sequence version replaced gi:1220812.
FEATURES             Location/Qualifiers
     source          1. .13299
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            complement(85. .1455)
                     /gene="HI0161"
     CDS             complement(85. .1455)
                     /gene="HI0161"
                     /note="similar to SP:P06715 GB:M13141 PID:146248
                     PID:466637 GB:U00096 percent identity: 74.17; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glutathione reductase (gor)"
                     /protein_id="AAC21833.1"
                     /db_xref="GI:1573119"
                     /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                     VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                     LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                     DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                     LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                     LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                     KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                     QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                     EFVTMR"
     gene            complement(1671. .2270)
                     /gene="HI0162"
     CDS             complement(1671. .2270)
                     /gene="HI0162"
                     /note="similar to GB:U00096 PID:1773118 PID:1786637
                     percent identity: 25.82; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="lipoprotein, putative"
                     /protein_id="AAC21834.1"
                     /db_xref="GI:1573120"
                     /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                     VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                     QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                     VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
     gene            2364. .2675
                     /gene="HI0163"
     CDS             2364. .2675
                     /gene="HI0163"
                     /note="similar to PID:663269 percent identity: 46.53;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, putative (bolA)"
                     /protein_id="AAC21835.1"
                     /db_xref="GI:1573121"
                     /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                     SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                     R"
     gene            2966. .4309
                     /gene="HI0164"
     CDS             2966. .4309
                     /gene="HI0164"
                     /note="similar to PID:663269 PID:663270 percent identity:
                     64.96; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreducatase subunit A, Na
                     translocating"
                     /protein_id="AAC21836.1"
                     /db_xref="GI:1573122"
                     /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                     VREGDVVKKGQVLFEDKKNPGVIFTAPASGTITAINRGEKRVLQSVVINVEGDEKITF
                     AKYSTEQLNTLSSEQVKQNLIESGLWTALRTRPFSKVPSIESEASSIFVNAMDTNPLA
                     ADPSVVLKEYSQDFTNGLTVLSRLFPSKPLHLCKAGDSNIPTADLENLQIHDFTGVHP
                     AGLVGTHIHFIDPVGIQKTVWHINYQDVIAVGKLFTTGELYSERVISLAGPQVKEPRL
                     VRTTIGANLSQLTQNELSAGKNRVISGSVLCGQIAKDSHDYLGRYALQVSVIAEGNEK
                     EFFGWIMPQANKYSVTRTVLGHFSKKLFNFTTSENGGERAMVPIGSYERVMPLDILPT
                     LLLRDLIVGDTDGAQELGCLELDEEDLALCSFVCPGKYEYGSILRQVLDKIEKEG"
     gene            4312. .5547
                     /gene="HI0166"
     CDS             4312. .5547
                     /gene="HI0166"
                     /note="similar to PID:663269 PID:663271 percent identity:
                     71.33; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, subunit B
                     (nqrB)"
                     /protein_id="AAC21837.1"
                     /db_xref="GI:1573123"
                     /translation="MGLKNLFEKMEPAFLPGGKYSKLYPIFESIYTLLYTPGTVTHKN
                     THVRDALDSKRMMITVFLALFPAIFYGMYNVGNQAIPALNQLGNLDQLIANDWHYALA
                     SSLGLDLTANATWGSKMALGAIFFLPIYLVVFTVCTIWELLFSVVRGHEVNEGMFVST
                     ILFALIVPPTLPLWQAALGITFGIIVAKEIFGGVGRNFMNPALAGRAFLFFAYPAQIS
                     GDTVWTAADGFSGATALSQWSQGGQGALQHTVTGAPITWMDAFVGNLPGSMGEVSTLA
                     ILIGGAVIVFTRIAAWRIIAGVMIGMIATSTLFNLIGSETNPMFSMPWHWHFVLGGFA
                     LGMVFMATDPVSASFTNTGKWWYGALIGVMAVLIRTVNPAYPEGMMLAILFANLFAPI
                     FDYIVVQANIKRRRARTNG"
     gene            5540. .6274
                     /gene="HI0167"
     CDS             5540. .6274
                     /gene="HI0167"
                     /note="similar to PID:663269 PID:663272 PID:893413 percent
                     identity: 54.36; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating"
                     /protein_id="AAC21838.1"
                     /db_xref="GI:1573124"
                     /translation="MAKFNKDSVGGTILVVLLLSLVCSIIVAGSAVMLKPAQEEQKLL
                     DKQKNILNVAGLLQENTNVKETYAKFIEPRFVDLATGEYTQQADDSQQAIPADADKAR
                     IRSRSKTTEVYLVKDEQGQTQQVILPIYGTGLWSVMYGLVSVQPDGNTINGITYYQHG
                     ETPGLGGEIENPNWASLFKGKKLFDEQHQPAIRIVKGQAPQDEHSIDGLSGATLTGNG
                     VQGTFNYWFSKDGFGPYLEKLHSGAN"
     gene            6274. .6900
                     /gene="HI0168"
     CDS             6274. .6900
                     /gene="HI0168"
                     /note="similar to PID:663269 PID:663273 PID:893414 percent
                     identity: 74.88; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating"
                     /protein_id="AAC21839.1"
                     /db_xref="GI:1573125"
                     /translation="MSGKTSYKDLLLAPIAKNNPIALQILGICSALAVTTKLETAFVM
                     AIAVTLVTGLSNLFVSLIRNYIPNSIRIIVQLAIIASLVIVVDQILKAYAYGLSKQLS
                     VFVGLIITNCIVMGRAEAFAMKSPPVESFVDGIGNGLGYGSMLIIVAFFRELIGSGKL
                     FGMTIFETIQNGGWYQANGLFLLAPSAFFIIGFVIWGLRTWKPEQQEK"
     gene            6904. .7500
                     /gene="HI0170"
```

```
     CDS             6904. .7500
                     /gene="HI0170"
                     /note="similar to PID:893416 PID:893415 percent identity:
                     84.34; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase, Na
                     translocating"
                     /protein_id="AAC21840.1"
                     /db_xref="GI:1573126"
                     /translation="MEHYISLFVKAVFIENMALSFFLGMCTFLAVSKKVSPAFGLGIA
                     VTFVLGIAVPVNQLIYANVLKENALIEGVDLSFLNFITFIGVIAGLVQILEMVLDKFM
                     PSLYNALGIFLPLIAVNCAIFGGVSFMVQRDYNFPESIVYGFGSGLGWMLAIVALAGL
                     TEKMKYADIPAGLKGLGITFISVGLMALGFMSFSGIQL"

   Query Match             40.9%;  Score 550.2;  DB 1;  Length 13299;
   Best Local Similarity   64.1%;  Pred. No. 6e-128;
   Matches  863;  Conservative    0;  Mismatches  478;  Indels    6;  Gaps    2;

Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| ||| || |||||| |  |||| || || ||||| | ||| |  |||||||
Db    2966 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 3025

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
           ||  ||||   ||  | |   |||| ||| | || || || || ||||| || |||||
Db    3026 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 3085

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           || || |||||  |   ||||| || |  || |||||||| ||||| || |||||||||
Db    3086 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTGAAGAC 3145

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
           ||||| ||||| || ||| | ||||| || || ||    || |  ||| | || ||  |
Db    3146 AAAAAAAATCCTGGTGTAATTTTTACAGCCCCTGCAAGCGGTACCATCACTGCAATCAAT 3205

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           |||||||||||| || ||| | || || || || |||   || |||||  | ||  |||||
Db    3206 CGTGGCGAAAAACGTGTATTACAATCTGTTGTTATTAATGTAGAAGGTGATGAAAAAATC 3265

Qy     301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
              || | |   ||     | |||    ||   | | |      ||| ||||   |
Db    3266 ACTTTTGCAAAATATAGCACAGAACAATTGAATACACTTTCTTCTGAACAAGTAAAACAA 3325

Qy     361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
           ||| ||||  |||| |||||||||||| ||  | || || ||||| || || ||| | ||
Db    3326 AACTTGATAGAATCTGGCTTATGGACGGCATTACGTACTCGTCCATTTAGTAAAGTTCCA 3385

Qy     421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           || | ||    ||  | |   | || || || |||||||||| ||||||||  | ||
Db    3386 TCCATTGAAAGTGAAGCATCTTCTATTTTTGTAAATGCGATGGATACCAATCCATTAGCA 3445

Qy     481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
           ||||| ||  | ||| |  | ||||||    | |||||||||    ||| |    |||
Db    3446 GCCGATCCCTCTGTCGTATTAAAAGAATATTCGCAAGACTTCACTAACGGTTTAACCGTA 3505

Qy     541 ttgagccgcctgaccgaacgtaa---aatccatgtgtgtaaagcagcaggcgcagacgtg 597
           || || || ||   |       ||   | | ||  | ||||||||||  |   |  |  |
Db    3506 TTAAGTCGTCTATTCCCTTCAAAACCATTACACTTATGTAAAGCAGGCGATTCTAATATT 3565

Qy     598 ccgtctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggc 657
           ||  |||   ||  ||| |  | |  ||||| |||   ||    ||||| || ||
Db    3566 CCAACTGCTGATCTTGAAAATTTACAGATTCATGATTTTACTGGTGTTCATCCAGCTGGT 3625

Qy     658 ttgagtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggacc 717
           |    || ||||| ||||| || || || || |||||     | ||||| || |||
Db    3626 CTTGTAGGGACGCATATTCACTTTATTGATCCTGTCGGTATTCAAAAAACTGTATGGCAT 3685

Qy     718 atcaattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaatacc 777
           |||||||||||||| ||||| ||| | ||    || ||   |||||||   ||  || |
Db    3686 ATCAATTATCAAGATGTGATCGCTGTAGGTAAATTATTTACAACAGGCGAACTTTATTCA 3745

Qy     778 gagcgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgtt 837
           || || ||  || |  | |  || |  ||||| ||  |||| ||  |  | |||||

Db    3746 GAACGAGTAATTTCTCTTGCAGGTCCTCAAGTGAAAGAACCTCGTTTAGTACGTACAACA 3805

Qy     838 ttgggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtg 897
           | ||||| ||  | |||||| | |||       |||||   ||  |   | || || ||
Db    3806 ATCGGTGCTAATCTTTCTCAATTAACCCAAAATGAATTAAGTGCAGGTAAAAATCGTGTA 3865

Qy     898 atttccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgc 957
           ||||| || || ||  |     |||   || ||| |||   | ||||||||||| || |||
Db    3866 ATTTCTGGCTCAGTTCTTTGTGGTCAAATAGCAAAAGATTCACATGATTATTTAGGTCGC 3925

Qy     958 taccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggtt 1017
           ||       ||  | || || ||||  |||||       |||||  | || || ||| ||
Db    3926 TATGCTTTGCAAGTATCAGTAATCGCTGAAGGTAATGAAAAAGAATTTTTTGGTTGGATT 3985

Qy    1018 gcgccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaac 1077
             ||| ||  |  | ||||| ||  | || || |||    | |||||  |||   |  ||
Db    3986 ATGCCACAAGCCAATAAATATTCTGTGACTCGTACCGTATTAGGCCACTTC---AGTAAA 4042

Qy    1078 aaactcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggc 1137
           ||| | || || || || || ||  | |  || || ||||| || || ||||| || |||||
Db    4043 AAATTGTTTAATTTTACTACTTCAGAAAATGGTGGCGAACGTGCAATGGTGCCAATCGGT 4102

Qy    1138 acttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatc 1197
           |  |||||||||||||||| |||||||  | ||||| ||  | || || |||||||||
Db    4103 AGCTATGAGCGCGTAATGCCATTGGACATTTTACCTACATTATTATTACGTGATTTAATC 4162

Qy    1198 gtcggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctc 1257
           || ||||||||||  | || || |  || |||||  ||||||| ||||||||||| |
Db    4163 GTGGGCGATACCGATGGTGCTCAAGAGTTAGGTTGTCTGGAATTAGACGAAGAAGACTTG 4222

Qy    1258 gctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtg 1317
           || || ||    ||||| || || |||||||| ||||||||  |  | |||||  ||||
Db    4223 GCGTTATGTTCTTTCGTTTGTCCAGGCAAATATGAATACGGTTCAATCTTGCGTCAAGTC 4282

Qy    1318 ctggaaaccattgagaaggaaggctga 1344
           | || |  |||||||||||||| | |
Db    4283 TTAGATAAGATTGAGAAGGAAGGTTAA 4309


RESULT  10
AF117331
LOCUS       AF117331                5916 bp    DNA     linear   BCT 13-MAY-1999
DEFINITION  Vibrio cholerae N16961 Na+-translocating NADH-ubiquinone
            oxidoreductase enzyme complex, complete sequence.
ACCESSION   AF117331
VERSION     AF117331.1  GI:4808841
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 5916)
  AUTHORS   Hase,C.C. and Mekalanos,J.J.
  TITLE     Effects of changes in membrane sodium flux on virulence gene
            expression in Vibrio cholerae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3183-3187 (1999)
  MEDLINE   99179036
REFERENCE   2  (bases 1 to 5916)
  AUTHORS   Hase,C.C. and Mekalanos,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1998) Microbiology and Molecular Genetics,
            Harvard Medical School, 200 Longwood Ave., Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .5916
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /db_xref="taxon:666"
                     /note="biotype: El Tor"
     gene            34. .1374
                     /gene="nqrA"
     CDS             34. .1374
                     /gene="nqrA"
                     /note="alpha subunit of NQR enzyme complex"
                     /codon_start=1
```

```
                /transl_table=11
                /product="NqrA"
                /protein_id="AAD29962.1"
                /db_xref="GI:4808842"
                /translation="MITIKKGLDLPIAGTPSQVISDGKAIKKVALLGEEYVGMRPTMH
                VRVGDEVKKAQILFEDKKNPGVKFTSPVSGKVVEINRGAKRVLQSVVIEVAGDDQVTF
                DKFEANQLASLNRDAIKTQLVESGLWTAFRTRPFSKVPAIDSTSEAIFVTAMDTNPLA
                AEPTVVINEQSEAFVAGLDVLSALTTGKVYVCKKGTSLPRSQQPNVEEHVFDGPHPAG
                LAGTHMHFLYPVSADHVAWSINYQDVIAVGQLFLTGELYTQRVVSLAGPVVNKPRLVR
                TVMGASLEQLVDSEIMPGEVRIISGSVLSGTKATGPHAYLGRYHLQVSVLREGRDKEL
                FGWAMPGKNKFSVTRSFLGHLFKGQVYNMTTTTNGSDRSMVPIGNYEKVMPLDMEPTL
                LLRDLCAGDSDSAVRLGALELDEEDLALCTFVCPGKYEYGQLLRECLDKIEKEG"
gene            1380. .2627
                /gene="nqrB"
CDS             1380. .2627
                /gene="nqrB"
                /note="a subunit of NQR enzyme complex"
                /codon_start=1
                /transl_table=11
                /product="NqrB"
                /protein_id="AAD29963.1"
                /db_xref="GI:4808843"
                /translation="MGLKKFLEDIEHHFEPGGKHEKWFALYEAAATLFYTPGLVTKRS
                SHVRDSVDLKRIMIMVWLAVFPAMFWGMYNAGGQALAALNHLYSGDQLAAIVAGNWHY
                WLTEMLGGTMSSDAGWGSKMLLGATYFLPIYATVFIVGGFWEVLFCMVRKHEVNEGFF
                VTSILFALIVPPTLPLWQAALGITFGVVVAKEVFGGTGRNFLNPALAGRAFLFFAYPA
                QISGDLVWTAADGYSGATALSQWAQGGAGALINNATGQTITWMDAFIGNIPGSIGEVS
                TLALMIGAAFIVYMGIASWRIIGGVMIGMILLSTLFNVIGSDTNAMFNMPWHWHLVLG
                GFAFGMFFMATDPVSASFTNSGKWAYGILIGVMCVLIRVVNPAYPEGMMLAILFANLF
                APLFDHVVVERNIKRRLARYGKQ"
gene            2614. .3387
                /gene="nqrC"
CDS             2614. .3387
                /gene="nqrC"
                /note="gamma subunit of NQR enzyme complex"
                /codon_start=1
                /transl_table=11
                /product="NqrC"
                /protein_id="AAD29964.1"
                /db_xref="GI:4808844"
                /translation="MASNNDSIKKTLFVVIALSLVCSIIVSAAAVGLRDKQKENAALD
                KQSKILQVAGIEAKGSKQIVELFNKSIEPRLVDFNTGDFVEGDAANYDQRKAAKEASE
                SIKLTAEQDKAKIQRRANVGVVYLVKDGDKTSKVILPVHGNGLWSMMYAFVAVETDGN
                TVSGLTYYEQGETPGLGGEVENPAWRAQWVGKKLFDENHKPAIKIVKGGAPQGSEHGV
                DGLSGATLTSNGVQNTFDFWLGDMGFGPFLTKVRDGGLN"
gene            3387. .4019
                /gene="nqrD"
CDS             3387. .4019
                /gene="nqrD"
                /note="b subunit of NQR enzyme complex"
                /codon_start=1
                /transl_table=11
                /product="NqrD"
                /protein_id="AAD29965.1"
                /db_xref="GI:4808845"
                /translation="MSSAKELKKSVLAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                TLAVMFVTALSNFFVSLIRNHIPNSVRIIVQMAIIASLVIVVDQILKAYLYDISKQLS
                VFVGLIITNCIVMGRAEAFAMKSEPIPSFIDGIGNGLGYGFVLMTVGFFRELLGSGKL
                FGLEVLPLISNGGWYQPNGLMLLAPSAFFLIGFMIWAIRTFKPEQVEAKE"
gene            4027. .4623
                /gene="nqrE"
CDS             4027. .4623
                /gene="nqrE"
                /note="c subunit of NQR enzyme complex"
                /codon_start=1
                /transl_table=11
                /product="NqrE"
                /protein_id="AAD29966.1"
                /db_xref="GI:4808846"
                /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGIA
                VIVVLTISVPVNNLVYNLVLKPDALVEGVDLSFLNFITFIGVIAALVQILEMILDRFF
                PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYSFAESVVYGFGSGVGWMLAIVALAGI
                REKMKYSDVPPGLRGLGITFITAGLMALGFMSFSGVQL"
gene            4657. .5883
                /gene="nqrF"
CDS             4657. .5883
                /gene="nqrF"
                /note="beta subunit of NQR enzyme complex"
                /codon_start=1
                /transl_table=11
                /product="NqrF"
                /protein_id="AAD29967.1"
                /db_xref="GI:4808847"
                /translation="MSTIIFGVVMFTLIILALVLVILFAKSKLVPTGDITISINGDPE
                KAIVTQPGGKLLTALAGAGVFVSSACGGGGSCGQCRVKIKSGGGDILPTELDHISKGE
                AREGERLACQVAVKADMDLELPEEIFGVKKWECTVISNDNKATFIKELKLAIPDGESV
                PFRAGGYIQIEAPAHHVKYADFDVPEKYRGDWDKFNLFRYESKVDEPIIRAYSMANYP
                EEFGIIMLNVRIATPPPNNPNVPPGQMSSYIWSLKAGDKCTISGPFGEFFAKDTDAEM
                VFIGGGAGMAPMRSHIFDQLKRLKSKRKMSYWYGARSKREMFYVEDFDGLAAENDNFV
                WHCALSDPQPEDNWTGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPMMNAAVINMLK
                NLGVEEEENILLDDFGG"
BASE COUNT     1441 a   1276 c   1550 g   1649 t
ORIGIN
```

```
  Query Match             38.6%;  Score 518.4;  DB 1;  Length 5916;
  Best Local Similarity   62.1%;  Pred. No. 6e-120;
  Matches  835;  Conservative    0;  Mismatches  506;  Indels    3;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| ||| ||||| ||  |  | || || ||||| || | ||    || || |||
Db     34 ATGATTACAATAAAAAAGGGATTGGACCTTCCTATCGCAGGAACTCCTTCCCAGGTGATT 93

Qy     61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
            ||| ||    ||||| |   ||||||| ||||||||||||||| || || ||||||||
Db     94 AGTGATGGTAAAGCCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 153

Qy    121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||  | ||| |  ||   | |||||| |  || |||||||  ||  | || ||||||||
Db    154 CCTACTATGCATGTCCGCGTAGGTGATGAAGTAAAAAAAGCTCAGATTCTTTTTGAAGAT 213

Qy    181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
          || |||||||||| ||   ||||||| | || |     || ||| | |  |  ||  |
Db    214 AAGAAGAATCCGGGTGTGAAATTTACTTCACCAGTCAGCGGTAAAGTGGTTGAAATCAAT 273

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          ||||| | |||||||||| |||||||| || |||||||  || |  ||  | ||| || |
Db    274 CGTGGTGCAAAGCGCGTTCTTCAGTCTGTAGTGATTGAAGTGGCGGGTGATGACCAAGTG 333

Qy    301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
             |||||    |  | | |  |     | || |  ||||   ||||     |
Db    334 ACCTTCGACAAGTTTGAAGCCAATCAATTAGCCAGTTTGAATCGCGATGCGATCAAAACT 393

Qy    361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
           | ||| || | || ||  | |||||||  | || || ||||||||||||||  | ||
Db    394 CAGCTGGTTGAGTCTGGTCTTTGGACTGCTTTCCGTACTCGTCCGTTCAGCAAGGTTCCA 453

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||| | ||| ||    |    ||||| || |||| ||| ||||| |||||||| ||||
Db    454 GCCATTGATTCCACTTCTGAAGCCATTTTTGTCACTGCTATGGATACCAATCCGTTGGCA 513

Qy    481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
          || || || || ||  | ||||| |||    | |||| ||||      || ||    ||
Db    514 GCTGAGCCGACCGTTGTGATCAATGAACAATCTGAAGCCTTCGTTGCTGGTCTAGATGTG 573

Qy    541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
            |   |   |||||     | ||| |  | || |||||    |  |||| |   | ||||
Db    574 CTTTCCGCACTGACAACTGGCAAAGTGTACGTTTGTAA---AAAAGGCACCAGCTTGCCA 630

Qy    601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
            |   |     | ||| | |||  |||||  |||| ||||||  |||||||| |||||
Db    631 CGTTCTCAACAGCCGAATGTTGAAGAACATGTCTTTGACGGCCCACATCCTGCAGGCTTA 690

Qy    661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
            |||||| ||||| |||||| || | |||||  | ||  |  |    | ||||| ||||
Db    691 GCTGGCACTCACATGCATTTCCTCTATCCAGTAAGTGCAGACCACGTGGCGTGGAGCATC 750
```

```
Qy    721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
          || ||||| |||||||| ||  | || |   |||| | || ||| , ||  | ||| ||
Db    751 AACTATCAGGACGTGATCGCAGTAGGTCAACTGTTCCTGACGGGCGAGCTATACACCCAG 810

Qy    781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
          || |||||| |  | | ||||| |   || |||||||| ||  |  | ||| || ||  ||
Db    811 CGTGTGGTTTCTCTTGCCGGCCCGGTGGTGAACAAACCTCGTTTGGTTCGCACTGTCATG 870

Qy    841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
          || ||     |    |||||    | | ||||| |  |    |  ||    ||| | ||
Db    871 GGCGCATCACTAGAGCAACTGGTGGACAGCGAAATCATGCCTGGCGAAGTACGCATCATC 930

Qy    901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
          || || || ||  ||   ||| |||  ||    ||  | || | |||| | || || |||
Db    931 TCTGGCTCTGTGCTGTCTGGTACGAAAGCCACTGGACCACACGCTTATCTTGGTCGTTAC 990

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
          |||   ||| | || ||| |    ||||| ||    ||||| ||||| || ||||    |
Db    991 CACCTACAGGTCTCTGTTCTTCGTGAAGGACGTGATAAAGAACTGTTTGGTTGGGCGATG 1050

Qy   1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
          |||     | |||| | |||  | || ||| |    || || ||  |  | ||    | ||
Db   1051 CCGGGTAAGAACAAGTTCTCTGTGACTCGCTCTTTCCTTGGTCACCTATTCAAGGGCCAA 1110

Qy   1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
           | | |||  | || ||| ||   || ||  | || ||  | ||||| || ||||| |
Db   1111 GTGTACAACATGACCACAACCACTAATGGTAGTGATCGATCTATGGTTCCAATCGGTAAC 1170

Qy   1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
          ||||||   || |||||  | || ||   ||| |||||| |  | || ||| |    |
Db   1171 TATGAGAAGGTCATGCCACTAGATATGGAGCCAACCTTGTTGCTTCGTGATCTGTGTGCT 1230

Qy   1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
          || || |  |||||||||      || |||   || ||  |||| |||||||| || ||
Db   1231 GGTGACAGTGACAGCGCGGTTCGCTTAGGTGCTTTAGAGCTGGATGAAGAAGATCTAGCA 1290

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
           |||| |  ||||| ||||| ||||||||||| ||||| | |||  ||||  |     |
Db   1291 CTGTGTACGTTCGTTTGCCCAGGCAAATACGAGTACGGTCAGCTCCTGCGTGAGTGCTTA 1350

Qy   1321 gaaaccattgagaaggaaggctga 1344
          || |  || ||||| ||||| | |
Db   1351 GACAAGATCGAGAAAGAAGGGTAA 1374
```

RESULT 11
AE004300/c
LOCUS AE004300 13382 bp DNA linear BCT 31-JUL-2000
DEFINITION Vibrio cholerae chromosome I, section 208 of 251 of the complete chromosome.
ACCESSION AE004300 AE003852
VERSION AE004300.1 GI:9656850
KEYWORDS .
SOURCE Vibrio cholerae.
ORGANISM Vibrio cholerae
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 13382)
AUTHORS Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L., Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D., Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I., Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D., Nierman,W.C. and White,O.
TITLE DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae
JOURNAL Nature 406 (6795), 477-483 (2000)
MEDLINE 20406833
REFERENCE 2 (bases 1 to 13382)
AUTHORS Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A., Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D., Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I., Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D., Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R., Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES Location/Qualifiers

```
source          1. .13382
                /organism="Vibrio cholerae"
                /strain="N16961"
                /serotype="O1"
                /db_xref="taxon:666"
                /chromosome="I"
                /note="biotype: El Tor"
gene            complement(154. .1236)
                /gene="VC2287"
CDS             complement(154. .1236)
                /gene="VC2287"
                /note="similar to GB:U00096 SP:Q47155 PID:1552799
                PID:1786425; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="DNA-damage-inducible protein P"
                /protein_id="AAF95431.1"
                /db_xref="GI:9656851"
                /translation="MQDRIRKIIHVDMDCFFAAVEMRDNPAYREIALAVGGHEKQRGV
                ISTCNYQARKFGVRSAMPTAQALKLCPQLHVVPGRMSVYKSVSQQIQTIFQRYTSLIE
                PLSLDEAYLDVSESTAYQGSATLIAQAIRRDIWQELNLTASAGVAPIKFLAKVASDLN
                KPDGLYVVTPDKVQEMVDSLPLEKIPGVGKVALEKLHQAGLYVGADVRRADYRKLLHQ
                FGRLGASLWKKSHGIDEREVVTERERKSVGVEYTFSQNISTFQECWQVIEQKLYPELD
                ARLSRAHPQRGIIKQGIKVKFADFQQTTIEHVHPALELDYFHELLEQVLTRQQGREIR
                LLGLSVMLKPELQMKQLSMFPSDGWQ"
gene            complement(1345. .1578)
                /gene="VC2288"
CDS             complement(1345. .1578)
                /gene="VC2288"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF95432.1"
                /db_xref="GI:9656852"
                /translation="MNTFLITFGFFLAVIAAMAVGYIFQRKVVKGSCGGLGAVGIEKV
                CNCPEPCDARKKREAKAAARAERLAAWEKDRIA"
gene            complement(1595. .2698)
                /gene="VC2289"
CDS             complement(1595. .2698)
                /gene="VC2289"
                /note="similar to GB:L42023 SP:P44550 PID:1003244
                PID:1222087 PID:1204428; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="thiamin biosynthesis lipoprotein ApbE"
                /protein_id="AAF95433.1"
                /db_xref="GI:9656853"
                /translation="MFFTNWILKVSMVIWFRVSVFPISIFNYERSKQVRNWLVALASL
                LLLAGCEKPAEQVHLSGPTMGTTYNIKYIQQPGIADSKTLQTEIDRLLEEVNDQMSTY
                RKDSELSRFNQHTSSEPFAVSTQTLTVVKEAIRLNGLTEGALDVTVGPLVNLWGFGPE
                ARPDVVPTDEELNARRAITGIEHLTIEGNTLSKDIPELYVDLSTIAKGWGVDVVADYL
                QSQGIENYMVEIGGEIRLKGLNRDGVPWRIAIEKPSVDQRSVQEIIEPGDYAIATSGD
                YRNYFEQDGVRYSHIIDPTTGRPINNRVVSVTVLDKSCMTADGLATGLMVMGEERGMA
                VAEANQIPVLMIVKTDDGFKEYASSSFKPFLSK"
gene            complement(2745. .3971)
                /gene="VC2290"
CDS             complement(2745. .3971)
                /gene="VC2290"
                /note="similar to PID:893416; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="NADH:ubiquinone oxidoreductase, Na
                translocating, beta subunit"
```

/protein_id="AAF95434.1"
/db_xref="GI:9656854"
/translation="MSTIIFGVVMFTLIILALVLVILFAKSKLVPTGDITISINGDPE
KAIVTQPGGKLLTALAGAGVFVSSACGGGGSCGQCRVKIKSGGGDILPTELDHISKGE
AREGERLACQVAVKADMDLELPEEIFGVKKWECTVISNDNKATFIKELKLAIPDGESV
PFRAGGYIQIEAPAHHVKYADFDVPEKYRGDWDKFNLFRYESKVDEPIIRAYSMANYP
EEFGIIMLNVRIATPPPNNPNVPPGQMSSYIWSLKAGDKCTISGPFGEFFAKDTDAEM
VFIGGGAGMAPMRSHIFDQLKRLKSKRKMSYWYGARSKREMFYVEDFDGLAAENDNFV
WHCALSDPQPEDNWTGYTGFIHNVLYENYLKDHEAPEDCEYYMCGPPMMNAAVINMLK
NLGVEEENILLDDFGG"

gene complement(4005. .4601)
/gene="VC2291"

CDS complement(4005. .4601)
/gene="VC2291"
/note="similar to PID:893416 PID:893415; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADH:ubiquinone oxidoreductase, Na
translocating, hydrophobic membrane protein NqrE"
/protein_id="AAF95435.1"
/db_xref="GI:9656855"
/translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGIA
VIVVLTISVPVNNLVYNLVLKPDALVEGVDLSFLNFITFIGVIAALVQILEMILDRFF
PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYSFAESVVYGFGSGVGWMLAIVALAGI
REKMKYSDVPPGLRGLGITFITAGLMALGFMSFSGVQL"

gene complement(4609. .5241)
/gene="VC2292"

CDS complement(4609. .5241)
/gene="VC2292"
/note="similar to PID:663269 PID:663273 PID:893414;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADH:ubiquinone oxidoreductase, Na
translocating, hydrophobic membrane protein NqrD"
/protein_id="AAF95436.1"
/db_xref="GI:9656856"
/translation="MSSAKELKKSVLAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
TLAVMFVTALSNFFVSLIRNHIPNSVRIIVQMAIIASLVIVVDQILKAYLYDISKQLS
VFVGLIITNCIVMGRAEAFAMKSEPIPSFIDGIGNGLGYGFVLMTVGFFRELLGSGKL
FGLEVLPLISNGGWYQPNGLMLLAPSAFFLIGFMIWAIRTFKPEQVEAKE"

gene complement(5241. .6014)
/gene="VC2293"

CDS complement(5241. .6014)
/gene="VC2293"
/note="similar to PID:663269 PID:663272 PID:893413;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADH:ubiquinone oxidoreductase, Na
translocating, gamma subunit"
/protein_id="AAF95437.1"
/db_xref="GI:9656857"
/translation="MASNNDSIKKTLFVVIALSLVCSIIVSAAAVGLRDKQKENAALD
KQSKILQVAGIEAKGSKQIVELFNKSIEPRLVDFNTGDFVEGDAANYDQRKAAKEASE
SIKLTAEQDKAKIQRRANVGVVYLVKDGDKTSKVILPVHGNGLWSMMYAFVAVETDGN
TVSGLTYYEQGETPGLGGEVENPAWRAQWVGKKLFDENHKPAIKIVKGGAPQGSEHGV
DGLSGATLTSNGVQNTFDFWLGDMGFGPFLTKVRDGGLN"

gene complement(6001. .7284)
/gene="VC2294"

CDS complement(6001. .7284)
/gene="VC2294"
/note="similar to PID:663269 PID:663271; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADH:ubiquinone oxidoreductase, Na
translocating, hydrophobic membrane protein NqrB"
/protein_id="AAF95438.1"
/db_xref="GI:9656858"
/translation="MLRQDRERRVISMGLKKFLEDIEHHFEPGGKHEKWFALYEAAAT
LFYTPGLVTKRSSHVRDSVDLKRIMIMVWLAVFPAMFWGMYNAGGQAIAALNHLYSGD
QLAAIVAGNWHYWLTEMLGGTMSSDAGWGSKMLLGATYFLPIYATVFIVGGFWEVLFC
MVRKHEVNEGFFVTSILFALIVPPTLPLWQAALGITFGVVVAKEVFGGTGRNFLNPAL
AGRAFLFFAYPAQISGDLVWTAADGYSGATALSQWAQGGAGALINNATGQTITWMDAF
IGNIPGSIGEVSTLALMIGAAFIVYMGIASWRIIGGVMIGMILLLSTLFNVIGSDTNAM
FNMPWHWHLVLGGFAFGMFFMATDPVSASFTNSGKWAYGILIGVMCVLIRVVNPAYPE
GMMLAILFANLFAPLFDHVVVERNIKRRLARYGKQ"

gene complement(7254. .8642)
/gene="VC2295"

CDS complement(7254. .8642)
/gene="VC2295"
/note="similar to PID:663269 PID:663270; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADH:ubiquinone oxidoreductase, Na
translocating, alpha subunit"
/protein_id="AAF95439.1"
/db_xref="GI:9656859"
/translation="MRNERIFFPIKVVQVRMITIKKGLDLPIAGTPSQVISDGKAIKK
VALLGEEYVGMRPTMHVRVGDEVKKAQILFEDKKNPGVKFTSPVSGKVVEINRGAKRV
LQSVVIEVAGDDQVTFDKFEANQLASLNRDAIKTQLVESGLWTAFRTRPFSKVPAIDS
TSEAIFVTAMDTNPLAAEPTVVINEQSEAFVAGLDVLSALTTGKVYVCKKGTSLPRSQ
QPNVEEHVFDGPHPAGLAGTHMHFLYPVSADHVAWSINYQDVIAVGQLFLTGELYTQR
VVSLAGPVVNKPRLVRTVMGASLEQLVDSEIMPGEVRIISGSVLSGTKATGPHAYLGR
YHLQVSVLREGRDKELFGWAMPGKNKFSVTRSFLGHLFKGQVYNMTTTTNGSDRSMVP
IGNYEKVMPLDMEPTLLLRDLCAGDSDSAVRLGALELDEEDLALCTFVCPGKYEYGQL
LRECLDKIEKEG"

gene complement(9079. .9399)
/gene="VC2296"

Query Match 38.6%; Score 518.4; DB 1; Length 13382;
Best Local Similarity 62.1%; Pred. No. 6.4e-120;
Matches 835; Conservative 0; Mismatches 506; Indels 3; Gaps 1;

```
Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| ||| ||||| ||  |  | || || ||||| || |  ||    || || |||
Db   8594 ATGATTACAATAAAAAAGGGATTGGACCTTCCTATCGCAGGAACTCCTTCCCAGGTGATT 8535

Qy     61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
            ||| ||    ||||| |   ||||||| ||||||||||||||| || || ||||||||
Db   8534 AGTGATGGTAAAGCCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 8475

Qy    121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||  | ||| |  ||   | |||||| |  || |||||||  ||  | || ||||||||
Db   8474 CCTACTATGCATGTCCGCGTAGGTGATGAAGTAAAAAAAGCTCAGATTCTTTTTGAAGAT 8415

Qy    181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
          || |||||||||||| ||   ||||||| | || |     || ||| | |  |  ||  |
Db   8414 AAGAAGAATCCGGGTGTGAAATTTACTTCACCAGTCAGCGGTAAAGTGGTTGAAATCAAT 8355

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
          ||||| | |||||||||| |||||||| || |||||||  || |  ||  | ||| || |
Db   8354 CGTGGTGCAAAGCGCGTTCTTCAGTCTGTAGTGATTGAAGTGGCGGGTGATGACCAAGTG 8295

Qy    301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
            |||||    |  | | |  |     | || |  ||||   ||||     |
Db   8294 ACCTTCGACAAGTTTGAAGCCAATCAATTAGCCAGTTTGAATCGCGATGCGATCAAAACT 8235

Qy    361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
          | ||| || | || ||  | ||||||||  | || || ||||||||||||||  | ||
Db   8234 CAGCTGGTTGAGTCTGGTCTTTGGACTGCTTTCCGTACTCGTCCGTTCAGCAAGGTTCCA 8175

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
          ||| | ||| ||    |    ||||| || |||| ||| ||||| |||||||| ||||
Db   8174 GCCATTGATTCCACTTCTGAAGCCATTTTTGTCACTGCTATGGATACCAATCCGTTGGCA 8115

Qy    481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
          || || || || ||  | ||||| |||    | |||| ||||       || ||    ||
Db   8114 GCTGAGCCGACCGTTGTGATCAATGAACAATCTGAAGCCTTCGTTGCTGGTCTAGATGTG 8055

Qy    541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
            |   | |||||      | ||| |  | || |||||   |  |||| |   | ||||
Db   8054 CTTTCCGCACTGACAACTGGCAAAGTGTACGTTTGTAA---AAAAGGCACCAGCTTGCCA 7998
```

```
Qy    601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
Db   7997 CGTTCTCAACAGCCGAATGTTGAAGAACATGTCTTTGACGGCCCACATCCTGCAGGCTTA 7938

Qy    661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
Db   7937 GCTGGCACTCACATGCATTTCCTCTATCCAGTAAGTGCAGACCACGTGGCGTGGAGCATC 7878

Qy    721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
Db   7877 AACTATCAGGACGTGATCGCAGTAGGTCAACTGTTCCTGACGGGCGAGCTATACACCCAG 7818

Qy    781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
Db   7817 CGTGTGGTTTCTCTTGCCGGCCCGGTGGTGAACAAACCTCGTTTGGTTCGCACTGTCATG 7758

Qy    841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
Db   7757 GGCGCATCACTAGAGCAACTGGTGGACAGCGAAATCATGCCTGGCGAAGTACGCATCATC 7698

Qy    901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
Db   7697 TCTGGCTCTGTGCTGTCTGGTACGAAAGCCACTGGACCACACGCTTATCTTGGTCGTTAC 7638

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db   7637 CACCTACAGGTCTCTGTTCTTCGTGAAGGACGTGATAAAGAACTGTTTGGTTGGGCGATG 7578

Qy   1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
Db   7577 CCGGGTAAGAACAAGTTCTCTGTGACTCGCTCTTTCCTTGGTCACCTATTCAAGGGCCAA 7518

Qy   1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
Db   7517 GTGTACAACATGACCACAACCACTAATGGTAGTGATCGATCTATGGTTCCAATCGGTAAC 7458

Qy   1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
Db   7457 TATGAGAAGGTCATGCCACTAGATATGGAGCCAACCTTGTTGCTTCGTGATCTGTGTGCT 7398

Qy   1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
Db   7397 GGTGACAGTGACAGCGCGGTTCGCTTAGGTGCTTTAGAGCTGGATGAAGAAGATCTAGCA 7338

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
Db   7337 CTGTGTACGTTCGTTTGCCCAGGCAAATACGAGTACGGTCAGCTCCTGCGTGAGTGCTTA 7278

Qy   1321 gaaaccattgagaaggaaggctga 1344
Db   7277 GACAAGATCGAGAAAGAAGGGTAA 7254
```

RESULT 12
AF165980
LOCUS AF165980 6048 bp DNA linear BCT 13-NOV-2001
DEFINITION Vibrio harveyi Na+-translocating NADH-quinone oxidoreductase complex operon, complete sequence.
ACCESSION AF165980
VERSION AF165980.1 GI:6531420
KEYWORDS .
SOURCE Vibrio harveyi.
ORGANISM Vibrio harveyi
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 6048)
AUTHORS Zhou,W., Bertsova,Y.V., Feng,B., Tsatsos,P., Verkhovskaya,M.L., Gennis,R.B., Bogachev,A.V. and Barquera,B.
TITLE Sequencing and preliminary characterization of the Na+-translocating NADH: ubiquinone oxidoreductase from Vibrio harveyi
JOURNAL Biochemistry 38 (49), 16246-16252 (1999)
MEDLINE 20056044
PUBMED 10587447
REFERENCE 2 (bases 1 to 6048)
AUTHORS Zhou,W., Feng,B., Gennis,R.B. and Barquera,B.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-1999) Biochemistry, University of Illinois, 600 South Mathews Avenue, Urbana, IL 61801, USA
FEATURES Location/Qualifiers
source 1. .6048
/organism="Vibrio harveyi"
/strain="BB120"
/db_xref="taxon:669"
gene 169. .1509
/gene="nqrA"
gene 169. .6022
/gene="Na+-translocating NADH-quinone oxidoreductase complex operon"
CDS 169. .1509
/gene="nqrA"
/codon_start=1
/transl_table=11
/product="NqrA"
/protein_id="AAF15411.1"
/db_xref="GI:6531421"
/translation="MITIKKGLALPIAGAPSQVINDGKTIKKVALLGEEYVGMRPTMH
VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEINRGAKRVLQSVVIEVAGEEQVTF
DKFEAAQLAGLDREVIKTQLVESGLWTALRTRPFSKVPAIESATKAIFVTAMDTNPLA
AQPELIITEQQEAFVAGLDILSALTEGKVYVCKSGTSLPSSSQSNVEEHVFDGPHPAG
LAGTHMHFLYPVNAENVAWSINYQDVIAFGQLFLTGELYTDRVVSLAGPVVNNPRLLR
TVVGASLEDLTDSELMPGEVRVISGSVLSGTQASGPHAYLGRYHQQVSVLREGRDKEL
FGWATPGKNKFSITKSFLGHIFKGQLFNMTTTTNGSDRAMVPIGNYERVMPLDMEPTL
LLRDLCAGDTDSAQTLGALELDEEDLALCTFVCPGKYEYGQLLRECLDTIVKEG"
gene 1513. .2754
/gene="nqrB"
CDS 1513. .2754
/gene="nqrB"
/codon_start=1
/transl_table=11
/product="NqrB"
/protein_id="AAF15412.1"
/db_xref="GI:6531422"
/translation="MLKKFIEDIEHHFEPGGKHEKWFALYEAAATLFYTPGLVTKRSS
HVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGDQLATVIAGNWHYW
LTEMLGGTIGAEAGVGSMMLLGATYFLPIYATVFIVGGFWEVLFCMVRKHEVNEGFFV
TSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPALAGRAFLFFAYPAQ
ISGDVVWTAADGFSGATALSQWAHRGSGALINNITGAPITWMDAFIGNIPGSIGEVST
LALMIGAAMIVYMRIASWRIIAGVMIGMIAVSTLFNVIGSDTNPMFNMPWHWHLVLGG
FAFGMFFMATDPVSASFTNSGKWWYGILIGAMCVMIRVVNPAYPEGMMLAILFANLFA
PLFDHVVIEKNIKRRLARYGK"
gene 2744. .3529
/gene="nqrC"
CDS 2744. .3529
/gene="nqrC"
/codon_start=1
/transl_table=11
/product="NqrC"
/protein_id="AAF15413.1"
/db_xref="GI:6531423"
/translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDQQKANAVLD
KQSKIVEVAGIDAEGKKVPELFAEYIEPRLVDFKTGDFVEKAEDGSTAANYDQRKAAK
DPAESIKLTADEDKAKILRRANTGIVYLVKNGDDISKVIIPVHGNGLWSMMYAFVAVE
TDGNTVSGITYYEQGETPGLGGEVENPVWRAQFVGKKLFDENHKPAIKIVKGGAPEGS
EHGVDGLSGATLTGNGVQGTFDFWLGDMGFGPFLAKVRDGGLN"
gene 3529. .4161
/gene="nqrD"
CDS 3529. .4161
/gene="nqrD"
/codon_start=1
/transl_table=11
/product="NqrD"
/protein_id="AAF15414.1"
/db_xref="GI:6531424"
/translation="MSSAQNIKKSIMAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
```
```

```
                     VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                     FGMEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
     gene            4169. .4765
                     /gene="nqrE"
     CDS             4169. .4765
                     /gene="nqrE"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrE"
                     /protein_id="AAF15415.1"
                     /db_xref="GI:6531425"
                     /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGVA
                     VVVVLTIAVPVNNLVYNLVLKENALVEGVDLSFLNFITFIGVIAALVQILEMVLDRFF
                     PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYNFAESIVYGFGSGVGWMLAIVALAGI
                     REKMKYSDVPPGLRGLGITFITVGLMALGFMSFSGVQL"
     gene            4799. .6022
                     /gene="nqrF"
     CDS             4799. .6022
                     /gene="nqrF"
                     /codon_start=1
                     /transl_table=11
                     /product="NqrF"
                     /protein_id="AAF15416.1"
                     /db_xref="GI:6531426"
                     /translation="MDIILGVVMFTLIVLALVLVILFAKSKLVPTGDITISVNDDPSL
                     AIVTQPGGKLLSALAGAGVFVSSACGGGGSCGQCRRKIKSGGGDILPTELDHITKGEA
                     REGERLACQVAMKTDMDIELPEEIFGVKKWECTVISNDNKATFIKELKLQIPDGESVP
                     FRAGGYIQIEAPAHHVKYADYDIPEEYREDWEKFNLFRYESKVNEETIRAYSMANYPE
                     EHGIIMLNVRIATPPPNNPDVAPGIMSSFIWSLKEGDKCTISGPFGEFFAKDTDAEMV
                     FVGGGAGMAPMRSHIFDQLKRLHSKRKMSFWYGARSKREMFYVEDFDGLAADNDNFVW
                     HCALSDPLPEDNWDGYTGFIHNVLYENYLRDHEAPEDCEYYMCGPPMMNAAVIGMLKD
                     LGVEDENILLDDFGG"
BASE COUNT     1571 a   1351 c   1455 g   1671 t
ORIGIN
```

```
  Query Match             38.2%;  Score 513.6;  DB 1;  Length 6048;
  Best Local Similarity   61.9%;  Pred. No. 9.8e-119;
  Matches  832;  Conservative    0;  Mismatches  509;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| ||| || || ||  |   ||| || ||||| ||    ||    || || |||
Db     169 ATGATTACAATAAAGAAGGGCTTGGCTCTTCCTATCGCAGGAGCTCCATCCCAGGTGATT 228

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
            |||| ||   | |||| |   ||||||| |||||||||||||| || || ||||||||
Db     229 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 288

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||  | ||| |  ||   |  || || |  || |||||||  ||||| || |||||||||
Db     289 CCTACCATGCATGTCCGCGTTGGCGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAC 348

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
           |||||||| || |||||    || ||||| || ||  | || ||| |   ||  ||| ||
Db     349 AAAAAGAACCCTGGCGTTAAGTTCACTGCTCCAGCAGCCGGTAAAGTGATCGAAATTAAC 408

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           |||||||  ||||| |||||||| ||||| |||||||  || | |||| | || | |  |
Db     409 CGTGGCGCTAAGCGTGTACTTCAATCAGTTGTGATTGAAGTAGCAGGCGAAGAGCAGGTT 468

Qy     301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
             ||| ||    | || | | |     || |||     |     |||||    |
Db     469 ACGTTTGATAAGTTCGAAGCAGCTCAACTAGCAGGTCTAGATCGCGAAGTGATCAAGACT 528

Qy     361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
            |  || || |||| ||| | ||||| ||  | || || ||||||||||||||  | ||
Db     529 CAATTGGTTGAATCTGGCCTTTGGACCGCTTTACGTACTCGTCCGTTCAGCAAGGTTCCA 588

Qy     421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
           ||  | ||  | |   |    || || ||||| | ||||||||| || ||||| || ||
Db     589 GCAATCGAGTCTGCTACTAAGGCTATTTTCGTAACTGCGATGGATACTAATCCACTAGCA 648

Qy     481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
```

```
           ||  | |||  | | |||||||  |||      ||||  |||      || ||     |
Db     649 GCACAGCCTGAGTTGATCATCACTGAACAACAAGAAGCATTCGTTGCAGGTCTAGACATT 708

Qy     541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
           |      |||||| ||| | ||  |  | || |||||     | ||| |   | ||||
Db     709 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAA---GTCTGGCACTAGCTTGCCA 765

Qy     601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
               |    | ||  | |||  ||| |  || |  || || || || || ||  |
Db     766 AGCTCTTCTCAGTCAAACGTAGAAGAACACGTCTTCGATGGTCCTCACCCAGCAGGTCTT 825

Qy     661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
             |||||| ||||| |||||| |  | |||||    ||  | ||    | ||||| ||||
Db     826 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCGTGGAGCATC 885

Qy     721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
           || || |||||||| || ||  |||| |   | ||| |||| ||    ||  | || ||
Db     886 AACTACCAAGACGTTATCGCGTTCGGTCAGCTTTTCCTAACGGGTGAGCTTTACACTGAT 945

Qy     781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
           || || ||  |  | |  || |     || ||||| || |||||  | ||||| ||| |
Db     946 CGTGTTGTGTCTCTAGCTGGTCCTGTTGTGAACAACCCACGCCTTCTTCGTACAGTTGTA 1005

Qy     841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
           |||||      |     | || || | | |||| ||| |    |  ||    || ||||||
Db    1006 GGTGCTTCTCTTGAAGACCTAACTGACAGCGAGTTGATGCCTGGTGAAGTTCGTGTGATT 1065

Qy     901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
           || || |||||| |    ||| |    ||    ||  | ||||  ||  | || || |||
Db    1066 TCAGGCTCGGTACTTTCTGGTACTCAAGCGTCTGGTCCACATGCCTACCTAGGTCGTTAC 1125

Qy     961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
           ||| | ||  |||| ||  |    ||||| ||   |||||| ||||| ||||||| | |
Db    1126 CACCAACAAGTTTCTGTGTTACGCGAAGGTCGTGACAAAGAACTGTTTGGCTGGGCTACA 1185

Qy    1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
           ||        |||| | ||| ||||||    |    || || ||  || | ||     |
Db    1186 CCTGGTAAAAACAAGTTCTCAATCACGAAATCTTTCCTTGGTCACATCTTCAAGGGTCAG 1245

Qy    1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
            | |||||  | || ||| |   ||||||  | ||||| || ||||| || |||||||
Db    1246 TTGTTCAACATGACAACAACGACCAACGGTAGTGACCGTGCAATGGTTCCAATCGGCAAC 1305

Qy    1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
           || |||||||||||||| | || ||    || || |||||  | || ||| ||   |
Db    1306 TACGAGCGCGTAATGCCGCTTGATATGGAACCGACTTTGCTACTTCGTGATCTATGTGCA 1365

Qy    1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
           |||||||| |||| || ||  | |||||    | ||| |||| |||||||| || ||
Db    1366 GGCGATACAGACAGTGCTCAAACACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 1425

Qy    1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
           ||||| | ||| || || || || |||||||| ||||| | |||  | ||  ||  ||
Db    1426 TTGTGTACCTTTGTATGTCCTGGTAAATACGAGTACGGTCAGCTACTTCGTGAATGCCTA 1485

Qy    1321 gaaaccattgagaaggaaggctga 1344
           || ||||| | |||||||||| | |
Db    1486 GATACCATCGTGAAGGAAGGGTAA 1509
```

```
RESULT  13
VANQRBOL
LOCUS       VANQRBOL                5078 bp    DNA     linear   BCT 15-JUN-1999
DEFINITION  V.alginolyticus bolA, nqrA, nqrB, nqrC, nqrD and nqrE (partial)
            genes.
ACCESSION   Z37111
VERSION     Z37111.1  GI:663268
KEYWORDS    bolA gene; BolA protein; NADH ubiquinone oxidoreductase;
            NADH:ubiquinone oxidoreductase subunit A; NADH:ubiquinone
            oxidoreductase subunit B; nqrA gene; nqrB gene; nqrC gene; nqrD
            gene; nqrE gene.
SOURCE      Vibrio alginolyticus.
  ORGANISM  Vibrio alginolyticus
```

```
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 2384)
  AUTHORS   Beattie,P., Tan,K., Bourne,R.M., Leach,D., Rich,P.R. and Ward,F.B.
  TITLE     Cloning and sequencing of four structural genes for the
            Na(+)-translocating NADH-ubiquinone oxidoreductase of Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 356 (2-3), 333-338 (1994)
  MEDLINE   95104445
REFERENCE   2  (bases 1 to 5078)
  AUTHORS   Ward,B.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1994) Ward B. F., Edinburgh University, Institute
            for Cell and Molecular Biology, Mayfield Rd, Edinburgh, Scotland,
            EH9 3JR
FEATURES             Location/Qualifiers
     source          1. .5078
                     /organism="Vibrio alginolyticus"
                     /db_xref="taxon:663"
                     /clone="10-2"
                     /clone_lib="Lambda library"
     gene            26. .334
                     /gene="bolA"
     CDS             26. .334
                     /gene="bolA"
                     /function="Cell division"
                     /codon_start=1
                     /transl_table=11
                     /product="BolA"
                     /protein_id="CAA85475.1"
                     /db_xref="GI:663269"
                     /db_xref="SWISS-PROT:Q56585"
                     /translation="MIQEIIEKKLHSELQPSYLKVINESYMHNVPPGSESHFKVIVVS
                     DSFAGQRLIGRHRQVHQILADELDNHIHALAIHTYTDEEWKREQDGAPDSPMCMGGGH
                     "
     gene            808. .2148
                     /gene="nqrA"
     CDS             808. .2148
                     /gene="nqrA"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="NADH:ubiquinone oxidoreductase subunit A"
                     /protein_id="CAA85476.1"
                     /db_xref="GI:663270"
                     /db_xref="SPTREMBL:Q56586"
                     /translation="MITIKKGLDLPIAGTPSQVINDGKTIKKVALLGEEYVGMRPTMH
                     VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEVNRGAKRVLQSVVIEVAGEEQVTF
                     DKFEAAQLSGLDREVIKTQLVDSGLWTALRTRPFSKVPAIESSTKAIFVTAMDTNPLA
                     AKPELIINEQQEAFIAGLDILSALTEGKVYVCKSGTSLPRSSQSNVEEHVFDGPHPAG
                     LAGTHMHFLYPVNAENVAWSINYQDVIAFGKLFLTGELYTDRVVSLAGPVVNNPRLVR
                     TVIGASLDDLTDNELMPGEVRVISGSVLTGTHATGPHAYLGRYHQQVSVLREGREKEL
                     FGWAMPGKNKFSVTRSFLGHVFKGQLFNMTTTTNGSDRSMVPIGNYERVMPLDMEPTL
                     LLRDLCAGDTDSAQALGALELDEEDLALCTFVCPGKYEYGTLLRECLDTIEKEG"
     gene            2118. .3398
                     /gene="nqrB"
     CDS             2118. .3398
                     /gene="nqrB"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase subunit B"
                     /protein_id="CAA85477.1"
                     /db_xref="GI:663271"
                     /db_xref="SPTREMBL:Q56587"
                     /translation="MPRYYREGRVIFMALKKFLEDIEHHFEPGGKHEKWFALYEAVAT
                     VFYTPGIVTNKSSHVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGD
                     QLATVISGNWHYWLTEMLGGTIAADAGVGSKMLLGATYFLPIYATVFLVGGFWEVLFC
                     MVRKHEVNEGFFVTSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPAL
                     AGRAFLFFAYPAQISGDVVWTAVDGFSGATALSQWAQGGNGALVNTVTGSPITWMDAF
                     IGNIPGSIGEVSTLALMIGAAMIVYMRIASGRIIAGVMIGMIAVSTLFNVIGSDTNPM
                     FNMPWHWHLVLGGFAFGMFFMATDPVSASFTNKGKWWYGILIGAMCVMIRVVNPAYPE
                     GMMLAILFANLFAPLFDHVVIEKNIKRRLARYGK"
     gene            3388. .4158
                     /gene="nqrC"
     CDS             3388. .4158
                     /gene="nqrC"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="NADH:uniquinone oxidoreductase"
                     /protein_id="CAA85478.1"
                     /db_xref="GI:663272"
                     /db_xref="SPTREMBL:Q56582"
                     /translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDKQKANAVLD
                     KQSKIVEVAGIDANGKKVPELFAEYIEPRLVDLETGNFTEGNASTYDQREASKDAERS
                     IALTPEEDVADIRRRANTAVVYLVKDQDEVQKVILPMHGKGLWSMMYAFVAVETDGNT
                     VSAITYYEQGETPGLGGEVENPSRRDQFIGKKLYNEDHQPAIKVVKGGAPQGSEHGVD
                     GLSGATLTSNGVQHTFDFWLGDKGFGPFLAKVRDGELN"
     gene            4158. .4790
                     /gene="nqrD"
     CDS             4158. .4790
                     /gene="nqrD"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:uniquinone oxidoreductase"
                     /protein_id="CAA85479.1"
                     /db_xref="GI:663273"
                     /db_xref="SPTREMBL:Q57095"
                     /translation="MSSAQNVKKSILAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                     FGLEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
     gene            4798. .5078
                     /gene="NqrE"
     CDS             4798. .>5078
                     /gene="NqrE"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH:ubiquinone oxidoreductase"
                     /protein_id="CAA85480.1"
                     /db_xref="GI:663274"
                     /db_xref="SWISS-PROT:Q56589"
                     /translation="MEHYISLLVKSISSKHALSFFLGMCTFLAVSKKVKTSFGLGVAV
                     VVVLTIAVPVNNLVYNLVLRENALVEGVDLSFLNFITFIGVIAALVQIL"
BASE COUNT     1366 a    1068 c    1207 g    1437 t
ORIGIN


  Query Match           36.3%;  Score 488;  DB 1;  Length 5078;
  Best Local Similarity   60.7%;  Pred. No. 2.8e-112;
  Matches  816;  Conservative    0;  Mismatches  525;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| ||| || || ||  |  |||| || ||||| || |  ||    || || |||
Db     808 ATGATTACAATAAAGAAGGGCTTGGATCTTCCTATCGCAGGAACTCCATCCCAGGTGATT 867

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
            |||| ||   | |||| |   ||||||| |||||||||||||| || || ||||||||
Db     868 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 927

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||  | ||| |  ||   |  ||||| |  || |||||||  ||||| || ||||||||
Db     928 CCTACCATGCATGTCCGCGTTGGTGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAT 987

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
           |||||||| || |||||   |||||||| || ||  | || ||| |   ||   || ||
Db     988 AAAAAGAACCCTGGCGTGAAATTTACTGCACCAGCAGCCGGTAAAGTGATCGAAGTTAAC 1047

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
           |||||||  || || || ||||| || || |||||||  || | |||  | ||  |  |
Db    1048 CGTGGCGCTAAACGTGTCCTTCAATCTGTAGTGATTGAAGTGGCAGGTGAAGAGCAGGTG 1107

Qy     301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
             || ||    | || | | |    ||  ||    |    | |||    |
Db    1108 ACATTTGATAAGTTCGAAGCCGCTCAACTTTCAGGTCTAGATCGTGAAGTGATCAAGACT 1167

Qy     361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
```

```
Db   1168 CAGTTGGTTGACTCTGGCCTATGGACCGCTTTACGTACTCGTCCGTTTAGCAAGGTTCCT 1227

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480

Db   1228 GCAATCGAGTCTTCTACTAAGGCTATTTTCGTAACTGCAATGGATACTAATCCATTAGCA 1287

Qy    481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540

Db   1288 GCTAAGCCTGAGTTGATCATTAATGAGCAACAAGAAGCATTCATTGCTGGTCTAGACATT 1347

Qy    541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600

Db   1348 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAA---ATCTGGCACTAGTTTGCCA 1404

Qy    601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660

Db   1405 CGCTCTTCACAGTCTAACGTAGAAGAACACGTCTTCGATGGCCCTCACCCAGCAGGTCTT 1464

Qy    661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720

Db   1465 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCATGGAGCATC 1524

Qy    721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780

Db   1525 AACTACCAAGACGTTATCGCGTTCGGTAAGCTGTTCCTAACTGGTGAGCTATACACAGAT 1584

Qy    781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840

Db   1585 CGTGTTGTGTCTCTAGCTGGTCCAGTTGTGAATAACCCTCGTCTAGTTCGTACCGTAATC 1644

Qy    841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900

Db   1645 GGTGCAAGCTTGGATGACCTAACAGATAACGAGTTGATGCCTGGCGAAGTTCGTGTGATT 1704

Qy    901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960

Db   1705 TCTGGTTCTGTACTGACAGGTACACATGCAACTGGTCCACACGCATACCTTGGCCGTTAC 1764

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020

Db   1765 CATCAACAGGTTTCTGTATTACGTGAAGGCCGTGAGAAAGAACTATTTGGCTGGGCAATG 1824

Qy   1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080

Db   1825 CCTGGTAAGAATAAGTTCTCGGTAACTCGCTCTTTCCTTGGTCACGTATTCAAAGGTCAG 1884

Qy   1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140

Db   1885 TTGTTCAATATGACAACGACGACCAATGGTAGTGATCGTTCAATGGTTCCAATTGGCAAC 1944

Qy   1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200

Db   1945 TACGAACGCGTAATGCCTCTAGATATGGAGCCTACTCTGCTGCTTCGTGATCTATGTGCA 2004

Qy   1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260

Db   2005 GGTGATACAGACAGTGCTCAAGCACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 2064

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320

Db   2065 TTGTGTACCTTTGTATGTCCAGGTAAGTACGAGTACGGTACGTTACTTCGTGAATGCCTA 2124

Qy   1321 gaaaccattgagaaggaaggctga 1344

Db   2125 GATACTATCGAGAAGGAAGGGTAA 2148


RESULT 14
AB008030
LOCUS       AB008030               6908 bp    DNA     circular BCT 15-JUN-1999
DEFINITION  Vibrio alginolyticus genes for Na-translocating NADH-quinone
            reductase complex, nqr operon, complete genome.
ACCESSION   AB008030 D43958 D49364
VERSION     AB008030.1  GI:2558472
KEYWORDS    Nqr1 subunit of Na-translocating NADH-quinone reductase complex
            alpha-subunit; Nqr2 subunit of Na-translocating NADH-quinone
            reductase complex; Nqr3 subunit of Na-translocating NADH-quinone
            reductase complex gamma-subunit; Nqr4 subunit of Na-translocating
            NADH-quinone reductase complex; Nqr5 subunit of Na-translocating
            NADH-quinone reductase complex; Nqr6 subunit of Na-translocating
            NADH-quinone reductase complex beta-subunit.
SOURCE      Vibrio alginolyticus DNA.
  ORGANISM  Vibrio alginolyticus
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (sites)
  AUTHORS   Hayashi,M., Hirai,K. and Unemoto,T.
  TITLE     Cloning of the Na(+)-translocating NADH-quinone reductase gene from
            the marine bacterium Vibrio alginolyticus and the expression of the
            beta-subunit in Escherichia coli
  JOURNAL   FEBS Lett. 356 (2-3), 330-332 (1994)
  MEDLINE   95104444
REFERENCE   2  (sites)
  AUTHORS   Beattie,P., Tan,K., Bourne,R.M., Leach,D., Rich,P.R. and Ward,F.B.
  TITLE     Cloning and sequencing of four structural genes for the
            Na(+)-translocating NADH-ubiquinone oxidoreductase of Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 356 (2-3), 333-338 (1994)
  MEDLINE   95104445
REFERENCE   3  (sites)
  AUTHORS   Hayashi,M., Hirai,K. and Unemoto,T.
  TITLE     Sequencing and the alignment of structural genes in the nqr operon
            encoding the Na(+)-translocating NADH-quinone reductase from Vibrio
            alginolyticus
  JOURNAL   FEBS Lett. 363 (1-2), 75-77 (1995)
  MEDLINE   95246889
REFERENCE   4  (sites)
  AUTHORS   Hayashi,M., Nakayama,Y. and Unemoto,T.
  TITLE     Existence of Na+-translocating NADH-quinone reductase in
            Haemophilus influenzae
  JOURNAL   FEBS Lett. 381 (3), 174-176 (1996)
  MEDLINE   96176316
REFERENCE   5  (sites)
  AUTHORS   Nakayama,Y., Hayashi,M. and Unemoto,T.
  TITLE     Identification of six subunits constituting Na+-translocating
            NADH-quinone reductase from the marine Vibrio alginolyticus
  JOURNAL   FEBS Lett. 422 (2), 240-242 (1998)
  MEDLINE   98149659
REFERENCE   6  (bases 1 to 6908)
  AUTHORS   Hayashi,M., Unemoto,T. and Sugiyama,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-1997) Maki Hayashi, Faculty of Pharmaceutical
            Sciences, Chiba University, Laboratory of Membrane Biochemistry;
            1-33 Yayoi-cho, Inage-ku, Chiba, Chiba 263, Japan
            (E-mail:makiha@p.chiba-u.ac.jp, Tel:043-290-2932, Fax:043-290-3021)
COMMENT     On Oct 22, 1997 this sequence version replaced gi:677955 gi:893412.
            D43958: submitted (17-Dec-1994)
            D49364: submitted (17-Feb-1995).
FEATURES             Location/Qualifiers
     source          1. .6908
                     /organism="Vibrio alginolyticus"
                     /db_xref="taxon:663"
                     /note="There are conflicts between this sequence and
                     GenBank Accession Number:Z37111."
     gene            386. .1726
                     /gene="nqr1"
     CDS             386. .1726
                     /gene="nqr1"
                     /note="contains FMN as a cofactor; nqrA, Nqr1 subunit of
                     Na-NQR complex alpha-subunit"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr1 subunit of Na-translocating NADH-quinone
                     reductase complex alpha-subunit"
                     /protein_id="BAA22910.1"
```

```
                     /db_xref="GI:2558473"
                     /translation="MITIKKGLDLPIAGTPSQVINDGKTIKKVALLGEEYVGMRPTMH
                     VRVGDEVKKAQVLFEDKKNPGVKFTAPAAGKVIEVNRGAKRVLQSVVIEVAGEEQVTF
                     DKFEAAQLSGLDREVIKTQLVDSGLWTALRTRPFSKVPAIESSTKAIFVTAMDTNPLA
                     AKPELIINEQQEAFIAGLDILSALTEGKVYVCKSGTSLPRSSQSNVEEHVFDGPHPAG
                     LAGTHMHFLYPVNAENVAWSINYQDVIAFGKLFLTGELYTDRVVSLAGPVVNNPRLVR
                     TVIGASLDDLTDNELMPGEVRVISGSVLTGTHATGPHAYLGRYHQQVSVLREGREKEL
                     FGWAMPGKNKFSVTRSFLGHVFKGQLFNMTTTTNGSDRSMVPIGNYERVMPLDMEPTL
                     LLRDLCAGDTDSAQALGALELDEEDLALCTFVCPGKYEYGTLLRECLDTIEKEG"
conflict             1205
                     /gene="nqr1"
                     /citation=[2]
                     /replace="c"
conflict             1234
                     /gene="nqr1"
                     /citation=[2]
                     /replace="g"
conflict             1256
                     /gene="nqr1"
                     /citation=[2]
                     /replace="t"
gene                 1732. .2976
                     /gene="nqr2"
CDS                  1732. .2976
                     /gene="nqr2"
                     /note="hydrophobic membrane protein with pI 8.14; nqrB,
                     Nqr2 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr2 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22911.1"
                     /db_xref="GI:2558474"
                     /translation="MALKKFLEDIEHHFEPGGKHEKWFALYEAVATVFYTPGIVTNKS
                     SHVRDSVDLKRIMIMVWFAVFPAMFWGMYNAGGQAIAALNHMYAGDQLATVISGNWHY
                     WLTEMLGGTIAADAGVGSKMLLGATYFLPIYATVFLVGGFWEVLFCMVRKHEVNEGFF
                     VTSILFALIVPPTLPLWQAALGITFGVVVAKEIFGGTGRNFLNPALAGRAFLFFAYPA
                     QISGDVVWTAADGFSGATALSQWAQGGNGALVNTVTGSPITWMDAFIGNIPGSIGEVS
                     TLALMIGAAMIVYMRIASWRIIAGVMIGMIAVSTLFNVIGSDTNPMFNMPWHWHLVLG
                     GFAFGMFFMATDPVSASFTNKGKWWYGILIGAMCVMIRVVNPAYPEGMMLAILFANLF
                     APLFDHVVIEKNIKRRLARYGK"
conflict             1881
                     /gene="nqr2"
                     /citation=[2]
                     /replace="c"
conflict             2400
                     /gene="nqr2"
                     /citation=[2]
                     /replace="t"
conflict             2417. .2418
                     /gene="nqr2"
                     /citation=[2]
                     /replace="tc"
conflict             2614
                     /gene="nqr2"
                     /citation=[2]
                     /replace="g"
gene                 2966. .3736
                     /gene="nqr3"
CDS                  2966. .3736
                     /gene="nqr3"
                     /note="acidic protein with pI 4.83; nqrC, Nqr3 subunit of
                     Na-NQR complex gamma-subunit"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr3 subunit of Na-translocating NADH-quinone
                     reductase complex gamma-subunit"
                     /protein_id="BAA22912.1"
                     /db_xref="GI:2558475"
                     /translation="MASNNDSIKKTLGVVIGLSLVCSIIVSTAAVGLRDKQKANAVLD
                     KQSKIVEVAGIDANGKKVPELFAEYIEPRLVDLETGNFTEGNASTYDQREASKDAERS
                     IALTPEEDVADIRRRANTAVVYLVKDQDEVQKVILPMHGKGLWSMMYAFVAVETDGNT
                     VSAITYYEQGETPGLGGEVENPSWRDQFIGKKLYNEDHQPAIKVVKGGAPQGSEHGVD
                     GLSGATLTSNGVQHTFDFWLGDKGFGPFLAKVRDGELN"
conflict             3295
                     /gene="nqr3"
                     /citation=[2]
                     /replace="c"
conflict             3514. .3515
                     /gene="nqr3"
                     /citation=[2]
                     /replace="ta"
gene                 3736. .4368
                     /gene="nqr4"
CDS                  3736. .4368
                     /gene="nqr4"
                     /note="hydrophobic membrane protein with pI 9.17; nqrD,
                     Nqr4 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr4 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22913.1"
                     /db_xref="GI:2558476"
                     /translation="MSSAQNVKKSILAPVLDNNPIALQVLGVCSALAVTTKLETAFVM
                     TLAVTFVTALSNFSVSLIRNHIPNSVRIIVQMAIIASLVIVVDQVLKAYLYDISKQLS
                     VFVGLIITNCIVMGRAEAFAMKSAPVPSLIDGIGNGLGYGFVLITVGFFRELFGSGKL
                     FGLEVLPLVSNGGWYQPNGLMLLAPSAFFLIGFLIWVIRILKPEQVEAKE"
gene                 4376. .4972
                     /gene="nqr5"
CDS                  4376. .4972
                     /gene="nqr5"
                     /note="hydrophobic membrane protein with pI 7.00; nqrE,
                     Nqr5 subunit of Na-NQR complex"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="Nqr5 subunit of Na-translocating NADH-quinone
                     reductase complex"
                     /protein_id="BAA22914.1"
                     /db_xref="GI:2558477"
                     /translation="MEHYISLLVKSIFIENMALSFFLGMCTFLAVSKKVKTSFGLGVA
                     VVVVLTIAVPVNNLVYNLVLRENALVEGVDLSFLNFITFIGVIAALVQILEMVLDRFF
                     PPLYNALGIFLPLITVNCAIFGGVSFMVQRDYNFAESIVYGFGSGVGWMLAIVALAGI
                     REKMKYSDVPPGLRGLGITFITVGLMALGFMSFSGVQL"
conflict             4410
                     /gene="nqr5"
                     /citation=[2]
                     /replace=""
conflict             4419
                     /gene="nqr5"
```

```
Query Match             36.0%;  Score 483.2;  DB 1;  Length 6908;
Best Local Similarity   60.5%;  Pred. No. 4.6e-111;
Matches  813;  Conservative    0;  Mismatches  528;  Indels    3;  Gaps    1;

Qy       1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
           ||||||| ||| || || ||  |  |||| || ||||| || |  ||    || || |||
Db     386 ATGATTACAATAAAGAAGGGCTTGGATCTTCCTATCGCAGGAACTCCATCCCAGGTGATT 445

Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
            |||| ||   | |||| |   ||||||| |||||||||||||| || || ||||||||
Db     446 AATGATGGTAAGACCATCAAAAAAGTCGCCTTGCTTGGCGAAGAGTACGTTGGCATGCGT 505

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           ||  | ||| |  ||   |  ||||| |  || |||||||  ||||| || ||||||||
Db     506 CCTACCATGCATGTCCGCGTTGGTGATGAAGTGAAAAAAGCGCAAGTTCTTTTTGAAGAT 565

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
           |||||||| || |||||   |||||||||| || ||  | || ||| |   ||   || ||
Db     566 AAAAAGAACCCTGGCGTGAAATTTACTGCACCAGCAGCCGGTAAAGTGATCGAAGTTAAC 625

Qy     241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
```

```
Db    626 CGTGGCGCTAAACGTGTCCTTCAATCTGTAGTGATTGAAGTGGCAGGTGAAGAGCAGGTG 685

Qy    301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360

Db    686 ACATTTGATAAGTTCGAAGCCGCTCAACTTTCAGGTCTAGATCGTGAAGTGATCAAGACT 745

Qy    361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420

Db    746 CAGTTGGTTGACTCTGGCCTATGGACCGCTTTACGTACTCGTCCGTTTAGCAAGGTTCCT 805

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480

Db    806 GCAATCGAGTCTTCTACTAAGGCTATTTTCGTAACTGCAATGGATACTAATCCATTAGCA 865

Qy    481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540

Db    866 GCTAAGCCTGAGTTGATCATTAATGAGCAACAAGAAGCATTCATTGCTGGTCTAGACATT 925

Qy    541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600

Db    926 CTTTCAGCCCTGACAGAAGGCAAGGTTTACGTATGTAA---ATCTGGCACTAGTTTGCCA 982

Qy    601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660

Db    983 CGCTCTTCACAGTCTAACGTAGAAGAACACGTCTTCGATGGCCCTCACCCAGCAGGTCTT 1042

Qy    661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720

Db   1043 GCTGGCACCCACATGCATTTCCTATACCCAGTGAATGCAGAAAATGTGGCATGGAGCATC 1102

Qy    721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780

Db   1103 AACTACCAAGACGTTATCGCGTTCGGTAAGCTGTTCCTAACTGGTGAGCTATACACAGAT 1162

Qy    781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgccttcttgcgtaccgttttg 840

Db   1163 CGTGTTGTGTCTCTAGCTGGTCCAGTTGTGAATAACCCTCGTTTAGTTCGTACCGTAATC 1222

Qy    841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900

Db   1223 GGTGCAAGCTTAGATGACCTAACAGATAACGAGCTGATGCCTGGCGAAGTTCGTGTGATT 1282

Qy    901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960

Db   1283 TCTGGTTCTGTACTGACAGGTACACATGCAACTGGTCCACACGCATACCTTGGCCGTTAC 1342

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020

Db   1343 CATCAACAGGTTTCTGTATTACGTGAAGGCCGTGAGAAAGAACTATTTGGCTGGGCAATG 1402

Qy   1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080

Db   1403 CCTGGTAAGAATAAGTTCTCGGTAACTCGCTCTTTCCTTGGTCACGTATTCAAAGGTCAG 1462

Qy   1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140

Db   1463 TTGTTCAATATGACAACGACGACCAATGGTAGTGATCGTTCAATGGTTCCAATTGGCAAC 1522

Qy   1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200

Db   1523 TACGAACGCGTAATGCCTCTAGATATGGAGCCTACTCTGCTGCTTCGTGATCTATGTGCA 1582

Qy   1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260

Db   1583 GGTGATACAGACAGTGCTCAAGCACTGGGTGCGCTTGAACTGGATGAAGAAGATCTAGCA 1642

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320

Db   1643 TTGTGTACCTTTGTATGTCCAGGTAAGTACGAGTACGGTACGTTACTTCGTGAATGCCTA 1702

Qy   1321 gaaaccattgagaaggaaggctga 1344
```

```
Db   1703 GATACTATCGAGAAGGAAGGGTAA 1726
```

RESULT 15
AE004724/c
LOCUS AE004724 16903 bp DNA linear BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 285 of 529 of the complete genome.
ACCESSION AE004724 AE004091
VERSION AE004724.1 GI:9949083
KEYWORDS .
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 16903)
AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P., Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
TITLE Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
JOURNAL Nature 406 (6799), 959-964 (2000)
MEDLINE 20437337
REFERENCE 2 (bases 1 to 16903)
AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA

FEATURES Location/Qualifiers
source 1. .16903
/organism="Pseudomonas aeruginosa"
/strain="PAO1"
/db_xref="taxon:287"
gene complement(67. .2292)
/gene="PA2984"
CDS complement(67. .2292)
/gene="PA2984"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06372.1"
/db_xref="GI:9949084"
/translation="MRLGLPALAAGLLALRFLAELPSVGWLLGMAAAGLLLLFGRLYW
LGLFLLGFAWACGSAQVALDGRLAAELDGRTLWLEGRVVDLPASGEGVTRFRLEEVSA
ARGVLLPERLRLSWRGAPPLMGGERWRLAVTLKRPRGLVNAAGFDYEAWLLAQRVGAV
GTVKAGQRLQPASGLDAWRDAWRQRLLAVDAQGRGAALAALVLGDGSGLSAAEWRMFQ
DTGTVHLMVISGSHISLLAGMLYGLVAALHRLGCWPRRLPWLPCACALGLLGAWSYGL
MAGFEVPVRRACLMVSLALVWRLRFRHLGLLTPLLATLVLVLLVEPLASLQAGFWLSF
LAVALLLWLFAGRLGRWRWWAAWGRAQWGMALGLAMPLLALGLPLSLSGAPANLLAVP
WIELLVVPPALAGSLLLWVPGLGEGLLRAAGVSLDLLFRALGLLAGLAEAWQPPAAAA
WSVALGMLGALCWLAPAGLPLRALGAALLLPALLPSSPPVEWGRAEVRVLDVGQGLAV
LVRTREHVLLYDSGARQGAFDMGERVVVPVLRSLDLRRLDGLLLSHADNDHAGGAPTV
ASRFPPVWLVSGEPARLPSPLFADSCDERSWSWDGVVFEQWAWAQAGDSNDRSCVLRV
EADGEVLLLTGDISRAAEHAWLARQADPRVDWLLAPHHGSRSSSGVAFVLRTRPRHVL
VSRGWRNAFGHPHGEVMERYARVAAQVHDTARDGALTFLLGSRGGARRERDSAHFWRE
K"
gene 2423. .2956
/gene="PA2985"
CDS 2423. .2956
/gene="PA2985"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06373.1"

```
                /db_xref="GI:9949085"
                /translation="MPRRLFKRYMPDPESIRSQKSLRFLGKLTHDPNLWHLNRHSVSR
                AMAAGLFAAFIPMPMQMLLAAGLAVWIRSNLPISVGLVWLTNPITMPPVFYCTYKLGA
                WLMQVPPRTLPDHLTWEWISQELSSLWQPFLLGSVVCGIIAAILLGYALTMLYWRWWVQ
                RSWSRRCRQRRERVAAK"
gene            complement(3044. .4345)
                /gene="PA2986"
CDS             complement(3044. .4345)
                /gene="PA2986"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG06374.1"
                /db_xref="GI:9949086"
                /translation="MPAVRSRPANPPSTRCIRDMFRPLPFFIGLRYTRAKRRNHYISF
                ISLTSMIGLALGVLAMIIVLSVMNGFQKEMRTRILGMVPHATISAAQPLDDWQTVANA
                ALRHREVVGAAPFAELQGMLSYKGNMLPVLVNGIDPQEERKVSIVGEHIVQGSLDDLK
                PGEFGIVLGEITARRFHVNVGDKLTLIVPEATSAPGGITPRMQRFTIVALFKVGAELD
                NSLALIDIADAGQLLRLQPGQVPSVRLELKDLYQSPQVAAKVVKELGQGFRSSDWTRT
                QGSLFNAMKMEKTMIGLLLLLIIAVAAFNIIATLIMVVADKRTDIAILRTLGATPRQI
                MAIFMVQGTVIGVIGTVIGGVLGVFAALNITGMIDRIERLVGHKVFSSDVYFINYLPS
                DLQVLDVVLICSAALLMSFLATLYPSWRAARTQPAESLRYE"
gene            complement(4358. .5041)
                /gene="PA2987"
CDS             complement(4358. .5041)
                /gene="PA2987"
                /codon_start=1
                /transl_table=11
                /product="probable ATP-binding component of ABC
                transporter"
                /protein_id="AAG06375.1"
                /db_xref="GI:9949087"
                /translation="MNDKSVLSCRNLSKSYDEGPQSVQVLSGVELNLLPGERVAIVGS
                SGSGKSTLLNMLGGLDTPSAGSVWLAGEELSALNETARGLLRNRALGFVYQFHHLLPE
                FTALENVCMPLLIGRTPIAEARQRAAELLERVGLGHRLSHKPAELSGGERQRVAIARA
                LVNRPQLVLLDEPTGNLDQHTAQGIQELMLELSRSLQTSFLVVTHDLQLAGHMDRILR
                LEEGRLIAA"
gene            complement(5034. .6284)
                /gene="PA2988"
CDS             complement(5034. .6284)
                /gene="PA2988"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG06376.1"
                /db_xref="GI:9949088"
                /translation="MFRPLSVFIGTRYTRAKRRSHFVSFISLTSMIGLALGVLVMIVV
                LSVMNGFDREMRTRILGMVPQATVESYQPIDDWRALAEKVKAHEHVTAVAPFTQMQGM
                LSANGAVQTVMVDAVDPLEEAKVSIIPDFFKQGSLAELKAGGFGIVIGQLAAQKLGVG
                IGDKVTLILPEVAVTPAGVFPRMKRFTVVGTFRVGAGELDGGLSLIHLEDAARLQRWK
                TNQVQGLRLKLDDLFQAPRVAWEIARTLTDNDFYARDWTRSHGNLYQAIRMEKTMIGL
                LLLLIVAVAAFNIISTLVMVVTDKKSDIAILRTLGATPGQIMATFMVQGTVIGVIGTL
                VGGVLGVVAALNVSAWISALEKLLGHQFLASDVYFIDYLPSQLMLDDVVLVCGAALVL
                SFFATLYPAWRAARTQPAEALRYE"
gene            6222. .6986
                /gene="PA2989"
CDS             6222. .6986
                /gene="PA2989"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG06377.1"
                /db_xref="GI:9949089"
                /translation="MAATFGPRIARADENGQGSEHVALSWQKMDASGWRSGEASPAAL
                HSDHHCRRGFAMSTRDADDRREYYRIEDTLALEFRPAGAVEDGLESEPGATLFSLLSD
                LHLMDYESQHLLRHISERDRTLANYLKVMNKRIDLLGQVMVQSLLKEIGEPRKVSLSE
                GGVSFRHDRALPVGQLLVLRMVLLPQGFGLELRARVIHAQPHDDEFEIGTEFEALSDA
                QRQLLARHILQKQAQQRRLARAGQGPLGEPGQPSST"
gene            7318. .8040
                /gene="PA2990"
CDS             7318. .8040
                /gene="PA2990"
                /codon_start=1
                /transl_table=11
                /product="probable phosphodiesterase"
                /protein_id="AAG06378.1"
                /db_xref="GI:9949090"
                /translation="MTLIYGHRGAKGEAPENTLASFQRCLEHGVNRCELDLHLSRDGE
                LMVIHDPTLKRTTGRRGKVAEHDAATLATYDARKGGPGWVRPTPIPRLEELFEKCPFE
                HWQLEVKSASRERAARTVTRIRELTERFGNRDKVTVTSGSRTVLKALNELTPELSRGL
                VAEYAWLDPLKVAQHYGCALLALNWTLCTPERLLKAQRQGLHVSVWTVNEPALMRRLA
                DFGVDSLITDFPGLASATLGNR"
gene            complement(8083. .9477)
                /gene="sth"
                /note="PA2991"
CDS             complement(8083. .9477)
                /gene="sth"
                /codon_start=1
                /transl_table=11
                /product="soluble pyridine nucleotide transhydrogenase"
                /protein_id="AAG06379.1"
                /db_xref="GI:9949091"
                /translation="MAVYNYDVVILGTGPAGEGAAMNASKYGRKLAVVDSRRVVGGNC
                THLGTIPSKALRHSVKQIIEFNTNPMFRQIGEPRWFSFPDVLKSADRVISKQVASRTG
                YYARNRIDMFTGTASFVDERTVEVVTPSGAVERLVADQFVIATGSRPYRPSDINFNHP
                RVYDSDTILSLSHTPRRLIIYGAGVIGCEYASIFSGLGVLVDLIDTRDQLLSFLDDEI
                SDALSYHLRNNNVLIRHNEEYERVEGLDNGVILHLKSGKKIKADALLWCNGRTGNTDK
                LGLENVGIKVNSRGQIEVDENYRTSVSNIFAAGDVIGWPSLASAAYDQGRSAAGNIVE
                SDSWRFVNDVPTGIYTIPEISSIGKNESELTAAKIPYEVGKAFFKGMARAQISNEPVG
                MLKILFHRETLEILGVHCFGDQASEIVHIGQAIMNQPGELNTLKYFVNTTFNYPTMAE
                AYRVAAFDGLNRLF"
gene            complement(9656. .9883)
                /gene="PA2992"
CDS             complement(9656. .9883)
                /gene="PA2992"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG06380.1"
                /db_xref="GI:9949092"
                /translation="MTWLLVFVIMLLVVFGMSIGVIMGRKPIAGSCGGIANLGIEKEC
                SICGGDRQKCEEANRDESVVPGKLAYDATKR"
gene            complement(9880. .10908)
                /gene="PA2993"
CDS             complement(9880. .10908)
                /gene="PA2993"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG06381.1"
                /db_xref="GI:9949093"
                /translation="MAERFRPLAMAGLAALALVLAGCGDTLESFGGPTMGSTYSIKYV
                RGASAPDVQTAKAAVEAILAEVDRQMSTYRDDSLVSRFNALPAQSCMELPLPMLELLR
                YGGELSEQSQGAFDMTVEPLMNLWGFGPQARVEKVPSAEQIAAVRRDVGHRHLRIDGQ
                RLCKDAAVQLDFDSIAAGYTVDAVGERLKELGVRSYLAEITGELKAEGRKPDGTPWRI
                AIEAPREGQRVAQQVLALDGYGVSTSGDYRNYFEENGQRYSHTFDPRTGAPIDHHLAS
                VTVIDPSTRNADGLSTLLMVLGPEEGYRFAEKHRLAALFVSRQGNGFDSRTTPRYEQL
                FGNQGERP"
gene            complement(10901. .12124)
                /gene="nqrF"
                /note="PA2994"
CDS             complement(10901. .12124)
```

```
  Query Match             35.4%;  Score 475.6;  DB 1;  Length 16903;
  Best Local Similarity   60.2%;  Pred. No. 4.1e-109;
  Matches  809;  Conservative    0;  Mismatches  529;  Indels    6;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||| || || |||   || ||  | |||||||||   |||   || || ||   |||
Db  16744 ATGATCAAGATAAAACGTGGCCTGGACCTGCCCATCAGCGGCGCCCCCGAACAGCGCATC 16685

Qy     61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
          | |  | |  |  | |   |  || |||||| | ||   || ||  |||||||
Db  16684 GAGGCAGCACGCCCGGTACGCAGCGTGGCGTTGATCGGTTTCGACTACCACGGCATGAAG 16625

Qy    121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
```

```
Db  16624 CCGACCATGGCGGTGCAGGTAGGCGACCGGGTGAAGCTGGGGCAGGTGCTGTTCACCGAC 16565

Qy    181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
Db  16564 AAGAAGAATCCCTCCGTCAGCTATACCGCACCCGGCGCCGGTGTGGTCAGTGCCATCCAT 16505

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
Db  16504 CGTGGCGAGAAGCGCGTCCTGCAATCGGTGGTCATCGATCTCGACGGCGACGAGCAACTG 16445

Qy    301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
Db  16444 GAGTTCGCCCGTTATCCCGCTGACAAACTCGCCACGCTGAGCGCCGAGCAGGTCCGCGAC 16385

Qy    361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
Db  16384 AACCTGTTGCAGTCGGGCCTGTGGACCGCGCTGCGGACCCGGCCGTTCAGCAAGGTGCCC 16325

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggct 480
Db  16324 GATCCCGAGTCCAGCCCCAGTTCGATTTTCGTCACTGCCATCGATACCCAGCCGCTGGCC 16265

Qy    481 gccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggta 540
Db  16264 GCCGATCCGCAGGTGGTTATCGCGGAGCAGGGCGAGGCCTTCCAGGCCGGGCTGACGGTG 16205

Qy    541 ttgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccg 600
Db  16204 CTCGGGCGCCTGGCC------CGGGTGTTCCTGTGCAAGGCGGAAGGCGTTTCGCTACCC 16151

Qy    601 tctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttg 660
Db  16150 GGCGAGGCGCTCTCCGGCGTGACCGCGCAAGCATTCTCCGGTCCGCATCCGGCGGGCCTG 16091

Qy    661 agtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatc 720
Db  16090 CCCGGCACCCATATCCACTTCCTTGACCCGGTGGGTGCCGGCAAGAGCGTGTGGAACCTG 16031

Qy    721 aattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgag 780
Db  16030 AACTACCAGGACGTGATCGCCATCGGCAAGCTGTTCACGACCGGGCAATTGTGGACCGAG 15971

Qy    781 cgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttg 840
Db  15970 CGGGTGATCGCCCTGGCCGGTCCGGTGGTGGAAAAGCCGCGTCTGGTGCGCACCCGCTTG 15911

Qy    841 ggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatt 900
Db  15910 GGCGCCAACCTCGACGAACTGGCCGCCGGCCAGTTGCAGCCGGGCAACAACCGGCTGATC 15851

Qy    901 tccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgggacgctac 960
Db  15850 TCCGGCTCCGTGCTGGGCGGTCGCACCGCCCATGGCGCCTACGCCTACCTGGGGCGCTAC 15791

Qy    961 cacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggttgcg 1020
Db  15790 CACCTGCAGTTGTCCTGTCTCAAGGAGGGCGACCAGCGCGAGTTCCTCCACTACCTGCGC 15731

Qy   1021 ccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaacaaa 1080
Db  15730 GCCGGGGTGGAAAAGCATTCGCTGCTCAACGTGTTCGTCTCCAGGCTGCTCGGCGGCAAG 15671

Qy   1081 ctcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcact 1140
Db  15670 CGCTTCGCCTTCACCACCAGCACCAACGGCAGCCCACGGGCGATGGTGCCGGTAGGTAAC 15611

Qy   1141 tatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtc 1200
Db  15610 TACGAGGCGGTGATGCCGCTGGATATCCTGCCGACTCAATTGCTGCGCTACCTGATCGTC 15551

Qy   1201 ggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgct 1260
```

```
Db  15550 GGCGACACGGAAATGGCCCAGAAGCTCGGCGCCCTGGAACTCGACGAAGAGGACCTGGCG 15491

Qy   1261 ttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctg 1320
Db  15490 CTGTGCAGCTACGTCTGCGCCGGCAAGTACGAATACGGCCCGATTCTGCGGGACAATCTG 15431

Qy   1321 gaaaccattgagaaggaaggctga 1344
Db  15430 GCACGCATCGAGCAGGAGGGCTGA 15407
```

RESULT 16
A96090
LOCUS A96090 474 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 123 from Patent WO9924578.
ACCESSION A96090
VERSION A96090.1 GI:6779937
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 474)
AUTHORS Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
TITLE Neisserial antigens
JOURNAL Patent: WO 9924578-A 123 20-MAY-1999;
PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES Location/Qualifiers
source 1. .474
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 122 a 130 c 130 g 91 t 1 others
ORIGIN

Query Match 32.3%; Score 433.6; DB 6; Length 474;
Best Local Similarity 94.5%; Pred. No. 1.2e-98;
Matches 448; Conservative 1; Mismatches 25; Indels 0; Gaps 0;

```
Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
Db      1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGCCGTT 60

Qy     61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
Db     61 TACGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGC 120

Qy    121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
Db    121 CCCTCGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180

Qy    181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
Db    181 AAAAAGAATCCGGGCGTGGTGTTTACTGCGCCGGCTTCAGGCAAAATCGCCGCGATTCAC 240

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatc 300
Db    241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAARGCAACGACGAAATC 300

Qy    301 gagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgc 360
Db    301 GAGTTTGAACGCTACGCACCTGAAGCGCTGGCAAACTTAAGCGGCGAAGAAGTGCGCCGC 360

Qy    361 aacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccct 420
Db    361 AACCTGATCCAATCCGGTTTGTGGACTGCGCTGCGCACCCGTCCGTTCAGCAAAATTCCT 420

Qy    421 gccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccg 474
Db    421 GCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCG 474
```

RESULT 17

```
AX067460/c
LOCUS       AX067460               96109 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 35 from Patent WO0078968.
ACCESSION   AX067460
VERSION     AX067460.1  GI:12545080
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 96109)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 35 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .96109
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
BASE COUNT    28783 a  18910 c  20341 g  28075 t
ORIGIN
```

Query Match 26.0%; Score 348.8; DB 6; Length 96109;
Best Local Similarity 55.2%; Pred. No. 4.6e-77;
Matches 749; Conservative 0; Mismatches 592; Indels 15; Gaps 3;

```
Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| |||||||||| | ||||||| ||| | |||   ||     |    |||
Db  85201 ATGATTACCATCAAAAAAGGCTTGGATCTGCCAATCACTGGCGAGCCCTCCAAGCAGATT 85142

Qy     61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
             ||   | || ||| ||   | || |||||  | ||| | || |||||||||||||||
Db  85141 AGCGAGCACACGCCCACTA---AGGTGGCGTTAGTCGGCTACGATTATGTCGGCATGCGT 85085

Qy    121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
          ||  | ||| |  |||| ||||||||    ||||  ||||| || |||||  | ||||||
Db  85084 CCGACCATGCATGTCAAAGAAGGTGATAAGGTCACCAAAGGACAGGTGCTTTACGAAGAC 85025

Qy    181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcac 240
          |||||     | |||||  | | ||| || || | |   ||||   |||  ||| | |
Db  85024 AAAAAACGAGCAGGCGTCCTTTATACAGCACCAGTTGGCGGCACTGTCGTTGCTGTCAAT 84965

Qy    241 cgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttg------aaggcaacgac 294
          |||||||||   || ||| || |     | || || ||||| |      | |  |  ||
Db  84964 CGTGGCGAACGACGAGTATTTGAAAGCTTGGTCATCGCCGTCGATCCAAATGCAAGTGAA 84905

Qy    295 gaaatcgagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtg 354
          ||||||   ||  |   ||     |||||||  || |||   ||    || | | |||
Db  84904 GAAATCACTTTTAATGCTTATCGCTCTGAAGATTTGTCAATGCTGGATCGCAAGACAGTT 84845

Qy    355 cgccgcaacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaa 414
                | || |    |||||   |||||| || || ||||| ||||| || |||
Db  84844 GTTGAGCAGTTGGTCGCTTCAGGTGAATGGACCGCCCTACGCACTCGTCCTTTTAGCCGT 84785

Qy    415 atccctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccg 474
          |  || |   |  ||  |   || |  ||||| || || |  || |  || ||||||||
Db  84784 ACACCAGAAATTAATAGCATACCTTCTGCCATTTTTGTGACGGCAACAGATACCAATCCT 84725

Qy    475 ctggctgccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctg 534
           ||||||  |||||| |   | |  |    ||||      |||  || ||    ||| ||
Db  84724 TTGGCTGTTGACCCTGCCACCGTATTATTTGAAGAAATACAAGCATTTAATGATGGCTTG 84665

Qy    535 ttggtattgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagac 594
              | ||      ||  ||  |   |          |     |     |   | ||
Db  84664 GCAATTTTATCAACGCTATCCCCAAAAACCTATGTCTGTCATGGTGATATTGTACCTGAG 84605

Qy    595 gtgccgtctgaaaatgctgccaatatcgaaac------acatgaatttggcggcccgcat 648
                | |||  |||| |  || ||| ||        | |  |||   ||    |||
Db  84604 ACCTTAAGTAAAACCGCTGACGGTAACGATACCATCTATGAAGGGTTTACTGGTGTACAT 84545

Qy    649 cctgccggcttgagtggcacgcacattcatttcatcgagccagtcggcgcgaataaaacc 708
          || || |||      || |||||||||| ||  |  | ||  | |       |
```

```
Db  84544 CCAGCAGGCAACGCAGGTACGCACATTCACTTTTTACACCCTTTGGCACGAGGCAGCTTA 84485

Qy    709 gtgtggaccatcaattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgt 768
          || ||||| || |||||||||| || || || || ||     | ||    ||||| ||
Db  84484 GTTTGGACGATTAATTATCAAGATGTCATCGCCATTGGTAAGCTATTTACCACAGGTCGC 84425

Qy    769 ctgaatacccgagcgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttg 828
            |  | || |||||  || |    ||||  || |    |||  ||| || ||| |  ||
Db  84424 ATTTACACAGAGCGTCTGATCAGTTTGGCAGGTCCAGCAGTGGCCAATCCTCGCTTGGTG 84365

Qy    829 cgtaccgttttgggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggac 888
          ||||||  |   |||||  |  || |  || | |||   || ||| |    | ||  ||
Db  84364 CGTACCACTCGAGGTGCTGATTTGACCGAATTGACCCAAGGTGAACTCAACGGCGACGAT 84305

Qy    889 aaccgcgtgatttccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattat 948
          ||||| || ||||| || || ||  |    |||   |  ||    || ||    | | |
Db  84304 AACCGTGTCATTTCAGGCTCTGTTCTATCAGGTCGCACAGCTGTTGGTGCTACGGCATTT 84245

Qy    949 ttgggacgctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttc 1008
          ||||| ||||  || |||||  | || ||  |    |||||  ||               |
Db  84244 TTGGGCCGCTTTCATAATCAAGTCTCTGTACTTACCGAAGGTCGTGAACGCCCTGCACTT 84185

Qy   1009 ggctgggttgcgccgcagccggacaaatactccatcacgcgcaccactctcggccatttc 1068
            ||   || | ||   | |  |    |  || |    ||  |    | |    || ||
Db  84184 CACTTTTTTACACCAGGGGCAAATCGTTTTTCAAAATTGCCAATTTATATTTCACAGTTT 84125

Qy   1069 ctaaaaaacaaactcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggta 1128
            | |||  ||||   | | | || |||||       |||||||          || |||||||||
Db  84124 TTTAAAGGCAAAAAATACGATTTTACGACCAGTACCAACGGCTCGCCTCGTGCCATGGTA 84065

Qy   1129 ccgatcggcacttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgc 1188
          || ||||||   | |||   || |||||    |||   || ||||||  |||  ||||
Db  84064 CCAATCGGCTTATTTGAAACTGTCATGCCTCAAGACTATCTACCTACCCAGCTACTGCGT 84005

Qy   1189 gatttaatcgtcggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaa 1248
          | ||| || ||||  |||| |  ||| |||    |  ||||||    | ||  | || |||
Db  84004 GCTTTGATTGTCGAAGATATCATCAGTGCGGTTGAATTGGGTGCACTTGAGCTTGATGAA 83945

Qy   1249 gaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttg 1308
          |||||  | ||  |||| |  ||||| |  || |||||||| || | |||     | |||
Db  83944 GAAGATTTGGCACTGTGTACTTTCGTATCGCCAGGCAAATATGAGTTCGGTGATATTTTG 83885

Qy   1309 cgcaaagtgctggaaaccattgagaaggaaggctga 1344
          ||  |     ||      |||||| |||||||| |
Db  83884 CGTGATAACTTGACCCGTATTGAGCAGGAAGGCTAA 83849
```

```
RESULT  18
AF144076
LOCUS       AF144076               1356 bp    DNA     linear   BCT 01-JUN-1999
DEFINITION  Porphyromonas gingivalis strain W50 50 kD antigen PG1 gene,
            complete cds.
ACCESSION   AF144076
VERSION     AF144076.1  GI:4929279
KEYWORDS    .
SOURCE      Porphyromonas gingivalis.
  ORGANISM  Porphyromonas gingivalis
            Bacteria; CFB group; Bacteroidetes; Bacteroidales;
            Porphyromonadaceae; Porphyromonas.
REFERENCE   1  (bases 1 to 1356)
  AUTHORS   Ross,B.C., Barr,I., Patterson,M., Agius,C., Rothel,L., Margetts,M.,
            Hocking,D. and Webb,E.
  TITLE     Porphyromonas gingivalis polypeptides and nucleic acids
  JOURNAL   Patent: Australia (AU 98/01023)-PCT 10-DEC-1998;
            CSL Ltd;
            45 Poplar Road;
            R&D;
            Parkville, Victoria;
            Australia;
REFERENCE   2  (bases 1 to 1356)
  AUTHORS   Ross,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1999) R&D, CSL Ltd, 45 Poplar Road, Parkville,
```

```
           Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .1356
                     /organism="Porphyromonas gingivalis"
                     /strain="W50"
                     /db_xref="taxon:837"
     CDS             1. .1356
                     /codon_start=1
                     /transl_table=11
                     /product="50 kD antigen PG1"
                     /protein_id="AAD33930.1"
                     /db_xref="GI:4929280"
                     /translation="MAKVIKTKKGLALNLKGKPLPEMLAEPAQSPTYAVVPDDFEGVI
                     PKVTARPGDKVRAGSALMHHKAYPEMKFTSPVSGEVIAVNRGAKRKVLSIEVKPDGLN
                     EYESFPVGDPSALSAEQIKELLLSSGMWGFIKQRPYDIVATPDIAPRDIYITANFTAP
                     LAPDFDFIVRGEERALQTAIDALAKLTTGKVYVGLKPGSSLGLHNAEIVEVHGPHPAG
                     NVGVLINHTKPINRGETVWTLKATDLIVIGRFLLTGKADFTRMIAMTGSDAAAHGYVR
                     IMPGCNVFASFPGRLTIKESHERVIDGNVLTGKKLCEKEPFLSARCDQITVIPEGDDV
                     DELFGWAAPRLDQYSMSRAYFSWLQGKNKEYVLDARIKGGERAMIMSNEYDRVFPMDI
                     YPEYLLKAIIAFDIDKMEDLGIYEVAPEDFATCEFVDTSKIELQRIVREGLDMLYKEM
                     N"
BASE COUNT      338 a    333 c    380 g    305 t
ORIGIN
```

Query Match 9.4%; Score 126.4; DB 1; Length 1356;
Best Local Similarity 46.7%; Pred. No. 4.4e-21;
Matches 631; Conservative 0; Mismatches 691; Indels 30; Gaps 6;

```
Qy       2 tgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcattt 61
Db      11 TTATAAAAACAAAAAAAGGCCTTGCACTTAATCTGAAAGGAAAACCGCTGCCCGAGATGC 70

Qy      62 atgacggcccggcc---attaccgaagtcgcgttgcttggcgaagaatatgtcggcatgc 118
Db      71 TGGCCGAACCGGCCCAAAGTCCTACTTACGCGGTCGTGCCCGACGATTTTGAAGGTGTTA 130

Qy     119 gcccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaag 178
Db     131 TCCCCAAGGTGACGGCTCGTCCGGGGGATAAGGTGCGTGCCGGCTCAGCACTGATGCACC 190

Qy     179 acaaaaagaatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattc 238
Db     191 ACAAGGCATATCCGGAGATGAAGTTTACAAGTCCGGTTAGCGGCGAAGTGATCGCGGTGA 250

Qy     239 accgtggcgaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaa 298
Db     251 ATCGCGGTGCCAAGCGCAAGGTGTTGAGCATCGAGGTGAAACCGGACGGACTGAACGAAT 310

Qy     299 tcgagttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgcc 358
Db     311 ACGAGTCATTCCCTGTCGGGGAT------CCGTCTGCCCTCTCTGCCGAACAGATCAAGG 364

Qy     359 gcaacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatcc 418
Db     365 AGCTTTTACTGTCGAGCGGTATGTGGGGTTTTATTAAGCAACGTCCTTAGGACATAGTGG 424

Qy     419 ctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctgg 478
Db     425 CTACACCGGATATAGCTCCACGCGACATTTATATTACTGCCAACTTTACTGCACCATTGG 484

Qy     479 ctgccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttgg 538
Db     485 CTCCGGACTTCGATTTCATCGTTCGAGGAGAAGAACGCGCCCTGCAGACTGCCATCGATG 544

Qy     539 tattgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgc 598
Db     545 CCTTGGCCAAACTCACGACAGGAAAGGTGTATGTGGGCCTGAAGCCGGGTTCATCTCTGG 604

Qy     599 cgtctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggct 658
Db     605 GCTTGCACAATGCAGAAATCGTAGAAGTACA---------CGGACCTCATCCGGCAGGTA 655

Qy     659 tgagtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggacca 718
Db     656 ACGTGGGCGTGCTGATCAATCATACGAAGCCAATCAATCGGGGCGAAACGGTGTGGACGC 715

Qy     719 tcaattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccg 778
Db     716 TCAAGGCTACCGACCTGATCGTGATCGGACGTTTCCTGCTTACGGGCAAAGCCGATTTTA 775

Qy     779 agcgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttt 838
Db     776 CCAGAATGATTGCCATGACCGGCTCAGACGCTGCAGCTCACGGATACGTCCGTATTATGC 835

Qy     839 tgggtgcgaaggtgtctcaacttaccgccggcgaattggtt------gacgcggacaacc 892
Db     836 CGGGTTGCAATGTCTTTGCTTCCTTCCCCGGCCGACTGACAATAAAGGAATCTCACGAGC 895

Qy     893 gcgtgatttccggttcggtattgaacggtgcgattgcacaaggcgcgcatgattatttgg 952
Db     896 GTGTGATCGATGGCAATGTGCTGACCGGTAAGAAGCTCTGCGAGAAGGAGCCTTTCCTGT 955

Qy     953 gacgctaccacaatcagatttccgttatcgaagaaggccgcagca---aagagctgttcg 1009
Db     956 CAGCCCGGTGTGACCAGATCACGGTGATCCCCGAAGGCGACGATGTGGACGAACTCTTCG 1015

Qy    1010 gctgggttgcgccgcagccggacaaatactccatcacgcgcaccactctcggccatttcc 1069
Db    1016 GGTGGGCTGCACCCCGTCTCGATCAGTACAGCATGAGCAGAGCTTATTTCTCTTGGTTGC 1075

Qy    1070 ---taaaaaacaaactcttcaagttcacgacagccgtcaacggcggcgaccgcgccatgg 1126
Db    1076 AGGGGAAAAACAAAGAGTACGTACTCGATGCCCGGATCAAGGGTGGCGAACGTGCTATGA 1135

Qy    1127 taccgatcggcacttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgc 1186
Db    1136 TCATGAGCAACGAGTATGACCGCGTTTTCCCGATGGACATCTATCCGGAGTATTTGCTCA 1195

Qy    1187 gcgatttaatcgtcggcgataccgacagcgcgcaggctttgggttgcttggaattggacg 1246
Db    1196 AGGCTATTATAGCATTCGACATCGACAAGATGGAGGACTTAGGCATATATGAAGTGGCTC 1255

Qy    1247 aagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgt 1306
Db    1256 CGGAGGACTTTGCCACTTGCGAATTTGTGGATACATCCAAGATCGAGCTGCAGCGTATCG 1315

Qy    1307 tgcgcaaagtgctggaaaccattgagaaggaa 1338
Db    1316 TTCGCGAGGGCTTGGATATGCTCTATAAGGAA 1347
```

```
RESULT 19
HIU20229/c
LOCUS       HIU20229               4926 bp    DNA     linear   BCT 09-FEB-1995
DEFINITION  Haemophilus influenzae BOLA (bolA), glutathione reductase (gor),
            phosphatidylserine decarboxylase (psd), 30K protein (rpmF), genes,
            complete cds.
ACCESSION   U20229
VERSION     U20229.1  GI:644850
KEYWORDS    .
SOURCE      Haemophilus influenzae.
  ORGANISM  Haemophilus influenzae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 4926)
  AUTHORS   Barcak,G.J. and Heimer,S.R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4926)
  AUTHORS   Barcak,G.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-1995) Gerard J. Barcak, Biochemistry, University
            of Maryland at Baltimore, 108 North Greene Street, Baltimore, MD
            21201, USA
FEATURES             Location/Qualifiers
     source          1. .4926
```

```
                     /organism="Haemophilus influenzae"
                     /strain="Rd"
                     /isolate="KW20"
                     /db_xref="taxon:727"
     CDS             complement(27. .224)
                     /note="orf65"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62134.1"
                     /db_xref="GI:644851"
                     /translation="MITIKKGLDLPIAGKPAQVIHSGNAVNQVAILGEEYVGMRPSMK
                     VREGDVVKKGQVLFEDKKILV"
     gene            complement(519. .842)
                     /gene="bolA"
     CDS             complement(519. .830)
                     /gene="bolA"
                     /function="stationary phase regulatory protein"
                     /codon_start=1
                     /transl_table=11
                     /product="BOLA"
                     /protein_id="AAA62135.1"
                     /db_xref="GI:644852"
                     /translation="MSIQQIIEQKIQKEFQPHFLAIENESHLHHSNRGSESHFKCVIV
                     SADFKNIRKVQRHQRIYQLLNEELNHSIHALALHLFTPEEWKAQNETVPHSTKCAGIG
                     R"
     misc_feature    775. .785
                     /note="encodes DNA uptake recognition sequence"
     RBS             complement(838. .842)
                     /gene="bolA"
     protein_bind    877. .913
                     /note="OxyR recongition sequence (Tartaglia, L.A., et al.,
                     J. Mol Biol. 210: 709-710, 1989)"
                     /bound_moiety="OxyR"
     RBS             915. .919
                     /note="orf199"
     CDS             924. .1523
                     /note="orf199"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62136.1"
                     /db_xref="GI:644853"
                     /translation="MALVNKIKTLSSVGILAATLFLAGCQAQSNILAFTPPAPSASMN
                     VNRTAVVSVTTKDSRAIQEIASYTKHGELIKLNASPSVTQLFQQVMQQNLISKGFRVG
                     QLNGSNAWVTVDVREFGTQVEQGNLRYKLNTKIQATVYVQGAKGSYNKSFNVTHSQEG
                     VFNAGNDEIHKVLSQTFNDIVNNIYQDQEVAAAINQYSN"
     gene            1343. .3109
                     /gene="gor"
     -35_signal      1343. .1348
                     /gene="gor"
                     /note="putative"
     -10_signal      1365. .1370
                     /gene="gor"
                     /note="putative"
     -35_signal      1585. .1590
                     /gene="gor"
                     /note="putative"
     -10_signal      1605. .1610
                     /gene="gor"
                     /note="putative"
     RBS             1728. .1731
                     /gene="gor"
     CDS             1739. .3109
                     /gene="gor"
                     /EC_number="1.6.4.2"
                     /function="reduces oxidized glutathione using NADPH"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="glutathione reductase"
                     /protein_id="AAA62137.1"
                     /db_xref="GI:644854"
                     /translation="MTKHYDYIAIGGGSGGIASLNRAASYGKKCAIIEAKHLGGTCVN
                     VGCVPKKVMFYGAHIAEAINNYAPDYGFDVEVKKFDFSKLIESRQAYISRIHTSYNNV
                     LAKNNIDVINGFGKFVDAHTIEVTLADGTKEQVTADHILIATGGRPYRPNIKGQEYGI
                     DSDGFFALTELPKRAAVIGAGYIAVELSGVLNSLGVETHLLVRRHAPMRNQDPLIVET
                     LVEVLAQDGIQLHTNSTPSEIVKNADGSLTVRCDGQSDVTVDCVIWAAGRVPTTDKIG
                     LENAGVETNEHGYVKVDKYQNTNVKGIYAVGDIIENGIELTPVAVAAGRRLSERLFNN
                     KPTEYLDYSLVPTVVFSHPPIGTVGLTEPQAIEQYGAENVKVYKSSFTAMYTAVTQHR
                     QPCKMKLVCVGKDEKVVGLHGIGFGVDEMIQGFAVAIKMGATKADFDNTVAIHPTGSE
                     EFVTMR"
     gene            complement(3218. .4090)
                     /gene="psd"
     CDS             complement(3218. .4090)
                     /gene="psd"
                     /function="lipid metabolism"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphatidylserine decarboxylase"
                     /protein_id="AAA62138.1"
                     /db_xref="GI:644855"
                     /translation="MMTSNSYWQRLKVAFQYVMPQIYLTQIAGWFAKQKWGKITHFVI
                     KAFAKKYNIDMSIAQKEQFSDYASFNEFFIRPLKENARPINQNPTALCLPADGRISEC
                     GHIDDNLLLQAKGHFFSLEDLLAEDKELVETFKNGEFVTTYLSPRDYHRVHMPCDATL
                     RKMIYVPGDLFSVNPFLAQHVPNLFARNERVICVFDTEFGTMVQILVGATITASIGTT
                     WAGVINPPRHNEVKTWTYEGESAVKLLKGQEMGWFQLGSTVINLFQANQVRLADHLSV
                     NEPVRMGEILAYKK"
     misc_feature    complement(3364. .3374)
                     /gene="psd"
                     /note="encodes DNA uptake recognition sequence"
     misc_feature    complement(3780. .3790)
                     /gene="psd"
                     /note="encodes DNA uptake recognition sequence"
     misc_feature    3818. .3825
                     /note="encodes DNA uptake recognition sequence"
     RBS             complement(4097. .4100)
                     /note="psd"
     protein_bind    complement(4126. .4152)
                     /note="OxyR recognition sequence (Toledano, M.B. et al.,
                     Cell 78:897-909.)"
                     /bound_moiety="OxyR"
     gene            4247. .4771
                     /gene="rpmF"
     CDS             4247. .4771
                     /gene="rpmF"
                     /codon_start=1
                     /transl_table=11
                     /product="30K protein"
                     /protein_id="AAA62139.1"
                     /db_xref="GI:644856"
                     /translation="MQKVKLPLTVDPIKDAQRRLDYVGYYAANQLERLAESVVNVLSD
                     AQVTLSFYVDPQKLVVMKGKVQIDVDLECQRCNEPYKQTLECEFTYSPVANWDQADDL
                     PEIYEPIEFNEFGEIDLIGTVEDELILALPLVPMHSSEHCEVSAQEQVFGVLPEELAK
                     KPNPFAVLANLKQK"
     CDS             complement(4561. .>4926)
                     /note="orf121"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA62140.1"
                     /db_xref="GI:644857"
                     /translation="LQLRGGVSEFHRLLLYLLIQQQLLKHRVNDAYHDVNEIFCFVEQP
                     LAILLNLIYFCFKLANTANGFGFFANSSGNTPKTCSCADTSQCSDECIGTSGKAKISS
                     SSTVPIKSISPNSLNSIGS"
BASE COUNT     1598 a    862 c    938 g   1528 t
ORIGIN


  Query Match             7.8%;  Score 105.4;  DB 1;  Length 4926;
  Best Local Similarity   70.0%;  Pred. No. 9.7e-16;
  Matches  156;  Conservative    0;  Mismatches   66;  Indels    1;  Gaps    1;

Qy      1 atgattaaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatt 60
          ||||||| ||| || |||||| |  |||| || || |||||| | ||| |  ||||| ||
Db    224 ATGATTACAATTAAGAAAGGTTTGGATCTCCCAATTGCGGGAAAACCAGCACAAGTAATC 165
```

```
Qy      61 tatgacggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgc 120
           ||  ||||   ||  | |   |||| ||| | || || || || ||||| || |||||
Db     164 CATAGCGGCAACGCTGTGAATCAAGTTGCGATTCTAGGTGAGGAGTATGTGGGGATGCGT 105

Qy     121 ccctcgatgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagac 180
           || || |||||  |    ||||| || |  || |||||||| ||||| || ||||||||||
Db     104 CCTTCAATGAAGGTGCGCGAAGGCGATGTTGTGAAAAAAGGTCAAGTACTTTTTGAAGAC 45

Qy     181 aaaaagaatccgggcgtagtatttactgcgccggcttcaggca 223
           ||||| ||||| || ||| | ||||| || || ||    || |
Db      44 AAAAA-AATCCTGGTGTAATTTTTACAGCCCCTGCAAGCGGTA 3
```

```
RESULT  20
BACGNTZA/c
LOCUS       BACGNTZA               36200 bp    DNA     linear   BCT 05-FEB-1999
DEFINITION  Bacillus subtilis 36kb sequence between gntZ and trnY genes
            encoding 34 ORFs.
ACCESSION   D78193
VERSION     D78193.1  GI:1064780
KEYWORDS    phospho-bete-glucosidase; 6-phosphogluconate dehydrogenase; alkyl
            hydroperoxide reductase; amino acid permease; arginase; orthinine
            aminotransferase; positive regulatory protein; gntZ; ahp; bglA;
            yyd; yyc; formaldehyde dehydrogenase.
SOURCE      Bacillus subtilis (strain:168) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 36200)
  AUTHORS   Ogasawara,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1995) Naotake Ogasawara, Nara Institute of
            Science and Technology, Graduate School of Biological Sciences;
            8916-5 Takayama-cho, Ikoma, Nara 630-01, Japan
            (E-mail:nogasawa@bs.aist-nara.ac.jp, Tel:07437-2-5430,
            Fax:07437-2-5439)
REFERENCE   2  (sites)
  AUTHORS   Zhang,J. and Aronson,A.
  TITLE     A Bacillus subtilis bglA gene encoding phospho-beta-glucosidase is
            inducible and closely linked to a NADH dehydrogenase-encoding gene
  JOURNAL   Gene 140 (1), 85-90 (1994)
  MEDLINE   94171085
REFERENCE   3  (sites)
  AUTHORS   Hartford,O.M. and Dowds,B.C.
  TITLE     Isolation and characterization of a hydrogen peroxide resistant
            mutant of Bacillus subtilis
  JOURNAL   Microbiology 140 (Pt 2), 297-304 (1994)
  MEDLINE   94236234
REFERENCE   4  (sites)
  AUTHORS   Ogasawara,N., Nakai,S. and Yoshikawa,H.
  TITLE     Systematic sequencing of the 180 kilobase region of the Bacillus
            subtilis chromosome containing the replication origin
  JOURNAL   DNA Res. 1 (1), 1-14 (1994)
  MEDLINE   96051385
REFERENCE   5  (sites)
  AUTHORS   Calogero,S., Gardan,R., Glaser,P., Schweizer,J., Rapoport,G. and
            Debarbouille,M.
  TITLE     RocR, a novel regulatory protein controlling arginine utilization
            in Bacillus subtilis, belongs to the NtrC/NifA family of
            transcriptional activators
  JOURNAL   J. Bacteriol. 176 (5), 1234-1241 (1994)
  MEDLINE   94156824
REFERENCE   6  (sites)
  AUTHORS   Yoshida,K., Seki,S., Fujimura,M., Miwa,Y. and Fujita,Y.
  TITLE     Cloning and sequencing of a 36-kb region of the Bacillus subtilis
            genome between the gnt and iol operons
  JOURNAL   DNA Res. 2 (2), 61-69 (1995)
  MEDLINE   96093926
REFERENCE   7  (sites)
  AUTHORS   Gardan,R., Rapoport,G. and Debarbouille,M.
  TITLE     Expression of the rocDEF operon involved in arginine catabolism in
            Bacillus subtilis
  JOURNAL   J. Mol. Biol. 249 (5), 843-856 (1995)
  MEDLINE   95311309
REFERENCE   8  (bases 1 to 36200)
  AUTHORS   Kasahara,Y., Nakai,S., Yoshikawa,H. and Ogasawara,N.
  TITLE     36kb sequence between gntZ and trnY of B. subtilis genome
  JOURNAL   Unpublished (1997)
FEATURES             Location/Qualifiers
     source          1. .36200
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            1. .32
                     /gene="gntZ"
     CDS             <1. .32
                     /gene="gntZ"
                     /citation=[6]
                     /codon_start=3
                     /transl_table=11
                     /product="6-phosphogluconate dehydrogenase"
                     /protein_id="BAA11267.1"
                     /db_xref="GI:1064781"
                     /translation="GVFHTNWSE"
     gene            496. .1059
                     /gene="ahpC"
     CDS             496. .1059
                     /gene="ahpC"
                     /note="induced by hydrogen peroxide"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="alkyl hydroperoxide reductase"
                     /protein_id="BAA11268.1"
                     /db_xref="GI:1064782"
                     /translation="MSLIGKEVLPFEAKAFKNGEFIDVTNEDLKGQWSVFCFYPADFS
                     FVCPTELEDLQEQYAALKELGVEVYSVSTDTHFVHKGWHDSSEKISKITYAMIGDPSQ
                     TISRNFDVLDEETGLADRGTFIIDPDGVIQTVEINAGGIGRDASNLVNKVKAAQYVRQ
                     NPGEVCPAKWEEGGETLTPSLDLVGKI"
     gene            1073. .2602
                     /gene="ahpF"
     CDS             1073. .2602
                     /gene="ahpF"
                     /note="induced by hydrogen peroxide"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="alkyl hydroperoxide reductase"
                     /protein_id="BAA11269.1"
                     /db_xref="GI:1064783"
                     /translation="MVLDANIKAQLNQYMQLIENDIVLKVSAGEDDTSKDMLALVDEL
                     ASMSSKISVEKAELNRTPSFSVNRVGEDTGVTFAGIPLGHEFTSLVLALLQVSGRPPK
                     VDQKVIDQVKKISGEYHFESYISLTCHNCPDVVQALNMMSVLNPNITHTMIDGAAYKA
                     EVESKNIMAVPTVYLNGESFGSGRMTLEEILAKMGSGTDASEFADKEPFDVLVVGGGP
                     AGASAAIYTARKGIRTGVVAERFGGQVLDTMSIENFISVKATEGPKLAASLEEHVKEY
                     DIDVMNLQRAKRLEKKDLFELELENGAVLKSKTVILSTGARWRNVNVPGEQEFKNKGV
                     AYCPHCDGPLFEGKDVAVIGGGNSGIEAAIDLAGIVNHVTVLEFAPELKADEVLQKRL
                     YSLPNVTVVKNAQTKEITGDQSVNGITYVDRETGEEKHVELQGVFVQIGLVPNTEWLE
                     GTVERNRMGEIIVDKHGATSVPGLFAAGDCTDSAYNQIIISMGSGATAALGAFDYLIR
                     N"
     gene            complement(2712. .4151)
                     /gene="bglA"
     CDS             complement(2712. .4151)
                     /gene="bglA"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="phospho-bete-glucosidase"
                     /protein_id="BAA11270.1"
                     /db_xref="GI:1064784"
                     /translation="MGNMPKDFLWGGALAAHQFEGGWNQGGKGPSVVDVMTAGAHGVP
                     RKITDTIEENEFYPNHEAIDFYHRYKEDIALFAEMGLKCLRTSIGWSRIFPKGDEAEP
                     NEAGLQFYDDVFDELLKHGIEPVITLSHFEMPLHLAREYGGFRNRKVVDFFVNFAEAC
                     FTRYKDKVKYWMTFNEINNQMDVNNPLFLWTNSGVVVGENENAKEVMYQTAHHELVAS
                     ALAVAKGKDINPEFQIGAMVSHVPIYPFSSNPEDVMLAEEEMRQRYFFPDVQVRGYYP
```

```
                SYALKEFEREGYNITFEDGDDEILRNGTVDYLGFSYYMSTTVKSDVKNDNTGDIVNGG
                LPNGVENPYITSSDWGWAIDPTGLRYTLNRFYDRYQIPLFIVENGFGAVDTLEEDGKV
                HDPERIQYLKSHIEALKKAVTYDGVDLIGYTPWGIIDIVSFTTGEMKKRYGMIYVDRD
                NEGNGSMKRYKKDSFEWYKNVIQTNGEEL"
     gene       4733. .5449
                /gene="yydK"
     CDS        4733. .5449
                /gene="yydK"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA11271.1"
                /db_xref="GI:1064785"
                /translation="MRMLKYQQIATEIETYIEEHQLQQGDKLPVLETLMAQFEVSKST
                ITKSLELLEQKGAIFQVRGSGIFVRKHKRKGYISLLSNQGFKKDLEDFNVTSKVIELD
                VRKPTPEAAENLNIGMDEDIYYVKRVRYINGQTLCYEESYYTKSIVTYLNNEIVSHSI
                FHYIREGLGLKIGFSDLFLHVGQLNEEEAEYLGLEAGLPKLYIESIFHLTNGQPFDYS
                KISYNYEQSQFVVQANSFLL"
     gene       complement(5766. .6488)
                /gene="yydJ"
     CDS        complement(5766. .6488)
                /gene="yydJ"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA11272.1"
                /db_xref="GI:1064786"
                /translation="MKLEFKKSISNKVIIILGAMFVFLFLLGYFLLVGIDKVSNVTPE
                MFFFSSYTVATQFGLMLFSFVIAFFINREYSNKNILFYKLIGENIYTFFYKKIAVLFL
                ECFAFITLGLLIISLMYHDFSHFALLLFLFSAVILQYILIIGTISVLCPNILISIGVS
                IVYWMTSVILVAISNKTFGFIAPFEAGNTMYPRIERVLQSDNMTLGSNDVLFIILYLV
                SIIIINAIVLRFSKTRWIKMGL"
     gene       complement(6509. .7138)
                /gene="yydI"
     CDS        complement(6509. .7138)
                /gene="yydI"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA11273.1"
                /db_xref="GI:1064787"
                /translation="MNIANYTLKVKGKTLLQDTDLHFSSGKINHVVGKNGVGKSQLAK
                DFLLNNSKRIGRDIRQNVSLISSSSNIPNDVSKDFLLHFLSKKFDAKMIDKIAYLLNL
                DNIDGKVLIKNLSDGQKQKLKLLSFLLEDKNIIVLDEITNSLDKKTVIEIHGFLNKYI
                QENPEKIIINITHDLSDLKAIEGDYYIFNHQEIQQYHSVDKLIEVYINE"
     gene       complement(7288. .8046)
                /gene="yydH"
     CDS        complement(7288. .8046)
                /gene="yydH"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA11274.1"
                /db_xref="GI:1064788"
                /translation="MKILKYEDGKYEVLVQNNVFIKDKKSGEYYKNSLNSLSDKQLLR
                FKMYKEKVSPKFFYLFLSFTALMFILNYIHLIKLQNGLSSVFYGWKMWIIIVIYFIMN
                IVLHELGHIYSLKFFGKNFDKVGFKLNFYVFPAFYVQLNETYMLSRNEKIIVHLFGLF
                INYLLINTLELINQFTFSSEALTMAFMLFSSTLLWNLIPILNSDGYKILLAFLSLDEY
                SRFKTNHWLVLTIQIIGIGLAVNSVVHWILYIVN"
     gene       complement(8027. .8986)
                /gene="yydG"
     CDS        complement(8027. .8986)
                /gene="yydG"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA11275.1"
                /db_xref="GI:1064789"
                /translation="MYNKTVSINLDSRCNASCDHCCFSSSPTSTTRMEKEYIRELVTE
                FAKNKTIQVISFTGGEVFLDYKFLKELMEIIKPYEKQITLISNGFWGLSKKKVQEYFH
                DMNSLNVIALTISYDEYHAPFVKSSSIKNILEHSRKYPDIDISLNMAVTKDKMSNHIL
                EELGDSILGVKITKFPMISVGAAKTRIKQENIHKFYSLEDEDSLHCPGYDIVYHHDGE
                IYPCCSPAIFETKITLREEYNQSFERTVEKLNSNLLLFILRKEGFKWFLNILKENNKI
                EEFDIPYEFSSICGVCGSLFNSAEKINYFYPYMEKYYNENFKV"
     gene       complement(9044. .9193)
```

Query Match 3.2%; Score 42.4; DB 1; Length 36200;
Best Local Similarity 64.0%; Pred. No. 8.9;

Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy      77 ttaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgcccctcgatgaaaatca 136
           || |  |  ||| | ||  |   |    || || ||||| ||     ||  ||||  | |
Db    4337 TTGCGCACATCGGGCTGGATACGGTGACATTTGACGGCACGCCTTTTTCCTTGAAGGTAA 4278

Qy     137 aggaaggtgaagccgtcaaaaaaggccaagtgctgtttga 176
           | ||||| ||  |||||||||||||| |||||||  ||||
Db    4277 AAGAAGGCGATACCGTCAAAAAAGGCGAAGTGCTTGTTGA 4238
```

```
RESULT  21
AP003456
LOCUS       AP003456               151329 bp    DNA     linear   HTG 29-MAR-2001
DEFINITION  Oryza sativa chromosome 6 clone P0542E10, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP003456
VERSION     AP003456.1  GI:13486925
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0542E10.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
            clone:P0542E10
  JOURNAL   Published Only in DataBase (2001) In press
REFERENCE   2  (bases 1 to 151329)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .151329
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="6"
                     /clone="P0542E10"
BASE COUNT    40598 a  33769 c  34863 g  41449 t    650 others
ORIGIN
```

Query Match 3.2%; Score 42.4; DB 2; Length 151329;
Best Local Similarity 46.1%; Pred. No. 9.9;
Matches 142; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

```
Qy    1036 tactccatcacgcgcaccactctcggccatttcctaaaaaacaaactcttcaagttcacg 1095
           | |  ||||     | ||      || ||  ||| | ||   ||| |||||||    ||
Db   25278 TCCCTCATCGTCTTCGCCCTCGCCGCCCTCTTCTTCAACCTCAACCTCTTCAGCGGCATC 25337

Qy    1096 acagccgtcaacggcggcgaccgcgccatggtaccgatcggcacttatgagcgcgtaatg 1155
           | | ||||  || |  |  | ||| ||| |   ||  |   |   ||    |   |
Db   25338 TCCGACGTCGGCGCCATCCTCGACCCCAAGGTCCGCGTCATCCTCTCCAAGGCGCTCTCG 25397

Qy    1156 ccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtcggcgataccgacagc 1215
           |  ||   |  | |     |||  || ||   |||    |    || || | ||| |  |
Db   25398 CTCTTCCTCCCCGTCATGTCCTACCTCTTCTCCGAGGCCAAGAACGCCGGTGCCGCCGCC 25457
```

```
Qy   1216 gcgcaggctttgggttgcttggaattggacgaagaagacctcgctttgtgcagcttcgtc 1275
          |    ||    ||| ||  ||  | |  |||    || ||| |  || |         |
Db  25458 TCCGCCGCCGGCGGTAGCACGGGCTCGCTCGAGCTCGAGCTCTCGCTGCGGGCGCGGCTG 25517

Qy   1276 tgcccgggcaaatacgaatacggcccgctgttgcgcaaagtgctggaaaccattgagaag 1335
          ||  |              || |  |||| | |||||   | ||||    |||  || |
Db  25518 ATCCTCGTGTGGATGCTCCTCGTCGAGCTGCTCCGCAAGAAGGTGGAGGAGATTCGGATG 25577

Qy   1336 gaaggctg 1343
          || |  ||
Db  25578 GAGGCATG 25585
```

RESULT 22
BSUB0021/c
LOCUS BSUB0021 215534 bp DNA linear BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 21 of 21): from 3999281 to 4214814.
ACCESSION Z99124 AL009126
VERSION Z99124.1 GI:2636442
KEYWORDS .
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus.
REFERENCE 1 (bases 1 to 215534)
AUTHORS Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G., Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S., Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C., Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V., Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J., Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D., Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E., Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A., Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J., Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R., Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F., Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y., Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P., Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A., Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H., Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P., Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M., Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M., Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P., Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M., Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T., Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J., Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B., Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K., Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A., Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A., Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T., Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K., Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and Danchin,A.
TITLE The complete genome sequence of the gram-positive bacterium Bacillus subtilis
JOURNAL Nature 390 (6657), 249-256 (1997)
MEDLINE 98044033
REFERENCE 2 (bases 1 to 215534)
AUTHORS Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

```
FEATURES             Location/Qualifiers
     source          1. .215534
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            complement(11. .820)
                     /gene="yxjF"
     terminator      complement(11. .37)
                     /gene="yxjF"
     CDS             complement(47. .820)
                     /gene="yxjF"
                     /function="unknown"
                     /note="similar to gluconate 5-dehydrogenase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15933.1"
                     /db_xref="GI:2636443"
                     /db_xref="SWISS-PROT:P42317"
                     /translation="MRKQVALVTGAAGGIRFEIAREFAREGASVIVSDLRPEACEKAA
                     SKLAEEGFDAAAIPYDVTKEAQVADTVNVIQKQYGRLDILVNNAGIQHVAPIEEFPTD
                     TFEQLIKVMLTAPFIAMKHVFPIMKKQQFGRIINIASVNGLVGFAGKSAYNSAKHGVI
                     GLTKVGALEGAPHGITVNALCPGYVDTQLVRNQLSDLSKTRNVPYDSVLEQVIFPLVP
                     QKRLLSVKEIADYAVFLASEKAKGVTGQAVVLDGGYTAQ"
     gene            complement(837. .1487)
                     /gene="yxjE"
     CDS             complement(837. .1487)
                     /gene="yxjE"
                     /function="unknown"
                     /note="similar to 3-oxoadipate CoA-transferase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15934.1"
                     /db_xref="GI:2636444"
                     /db_xref="SWISS-PROT:P42316"
                     /translation="MKEARKRMVKRAVQEIKDGMNVNLGIGMPTLVANEIPDGVHVML
                     QSENGLLGIGPYPLEGTEDADLINAGKETITEVTGASYFDSAESFAMIRGGHIDLAIL
                     GGMEVSEQGDLANWMIPGKMVKGMGGAMDLVNGAKRIVVIMEHVNKHGESKVKKTCSL
                     PLTGQKVVHRLITDLAVFDFVNGRMTLTELQDGVTIEEVYEKTEADFAVSQSVLNS"
     gene            complement(1484. .2200)
                     /gene="yxjD"
     CDS             complement(1484. .2200)
                     /gene="yxjD"
                     /function="unknown"
                     /note="similar to 3-oxoadipate CoA-transferase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15935.1"
                     /db_xref="GI:2636445"
                     /db_xref="SWISS-PROT:P42315"
                     /translation="MGKVLSSSKEAAKLIHDGDTLIAGGFGLCGIPEQLILSIRDQGV
                     KDLTVVSNNCGVDDWGLGLLLANKQIKKMIASYVGENKIFERQFLSGELEVELVPQGT
                     LAERIRAGGAGIPGFYTATGVGTSIAEGKEHKTFGGRTYVLERGITGDVAIVKAWKAD
                     TMGNLIFRKTARNFNPIAAMAGKITIAEAEEIVEAGELDPDHIHTPGIYVQHVVLGAS
                     QEKRIEKRTVQQASGKGEAK"
     gene            complement(2279. .3643)
                     /gene="yxjC"
     CDS             complement(2279. .3643)
                     /gene="yxjC"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB15936.1"
                     /db_xref="GI:2636446"
                     /db_xref="SWISS-PROT:P42314"
                     /translation="MELIIILLALGLLMFTAYRGFSVILFAPICALFAVLLTDPSHVL
                     PFFSSIFMEKMAGFIKLYFPVFLLGAIFGKVVEMAGLAASIAKTIVRLVGAKRAILAI
                     VLMGAVLTYSGVSLFVVVFAVYPFAKNMFQEANIPKRLIPGTIALGAFTFTMDALPGT
                     PQIQNVIPTSFFKTDIYAAPWLGLMGAVIVLAAGMLYLESRRKKAQASGEGYGGFDSQ
                     NAPAPESIESAAEPDKSPIRHALAFVPLILVGAVNKYFTIYLPKWYPNGFNFPSIGLK
                     EFGRLDISSAAAIWSVEIALVIGIITTILFDWRSVFAQLKEGLNEGIGGALLASMNTG
                     AEYGFGGIIAALPGFHKLSSGISHTFTDPLVNGAVTTTALAGITGSASGGMGIALSAM
                     SEQYLQAIQAYNIPPEVMHRVISMASGGMDTLPHNGAVITLSGRDGFDPPAILSRYFC
                     DHAH"
     gene            complement(3762. .4645)
                     /gene="yxjB"
     terminator      complement(3762. .3788)
```

```
                /gene="yxjB"
     CDS        complement(3797. .4645)
                /gene="yxjB"
                /function="unknown"
                /note="similar to hypothetical proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15937.1"
                /db_xref="GI:2636447"
                /db_xref="SWISS-PROT:P42313"
                /translation="MKRTVDFSMFRCPLCDSSMDAASGKSLICTERGHTFDLSRHGYV
                NFLTKPVKTSYGAELFEARSRLIGECGFFDPLHDAIAELISHPKSGHEAFTILDSGCG
                EGSHLNALCGFDYAGKAAIGTGIDLSKDGILKASKAFKDLMWAVADVARAPFHDRQFD
                VVLSIFSPSNYAEFHRLLKNDGMLIKVVPRSDYLIELRQFLYTDSPRRTYSNTAAVER
                FTANAAHSRPVRLRYVKTLDQQAIHWLLKMTPLAWSAPKDRVSLLKEMKSADITVDVD
                ILIGMK"
     gene       5261. .6489
                /gene="yxjA"
     CDS        5261. .6454
                /gene="yxjA"
                /function="unknown"
                /note="similar to pyrimidine nucleoside transport"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15938.1"
                /db_xref="GI:2636448"
                /db_xref="SWISS-PROT:P42312"
                /translation="MYFLLNLVGLIVIMAVVFLCSPQKKKIKWRPIITLIVLELLITW
                FMLGTKVGSWAIGKIGDFFTWLIACASDGIAFAFPSVMANETVDFFFSALLPIIFIVT
                FFDILTYFGILPWLIDKIGWVISKASRLPKLESFFSIQMMFLGNTEALAVIRQQLTVL
                SNNRLLTFGLMSMSSISGSIIGSYLSMVPATYVFTAIPLNCLNALIIANLLNPVHVPE
                DEDIIYTPPKEEKKDFFSTISNSMLVGMNMVIVILAMVIGYVALTSAVNGILGVFVHG
                LTNQTIFAYLFSPFAFLLGLPVHDAMYVAQLMGMKLATNEFVAMLDLKNNLKSLPPHT
                VAVATTFLTSFANFSTVGMIYGTYNSILDGEKSTVIGRNVWKLLVSGIAVSLLSAAIV
                GLFVW"
     terminator 6462. .6489
                /gene="yxjA"
     gene       complement(6925. .7167)
                /gene="yxiT"
     CDS        complement(6925. .7167)
                /gene="yxiT"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15939.1"
                /db_xref="GI:2636449"
                /db_xref="SWISS-PROT:P42311"
                /translation="MKWNNMLKAAGIAVLLFSVFAYAAPSLKAVQAKTPTVSTHTYKK
                IKELTYPQVHHVGNAAFEKKINQELKAYMDNRIRNT"
     terminator complement(7290. .7306)
                /gene="yxiS"
     gene       complement(7290. .7602)
                /gene="yxiS"
     CDS        complement(7312. .7602)
                /gene="yxiS"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB15940.1"
                /db_xref="GI:2636450"
                /db_xref="SWISS-PROT:P42310"
                /translation="MNRKQLTEAYERDENMMILVFAQWCVNHDLDPMELYAKAYPQQK
                LNESLKKTMDDLVVPKHEAEHIPDQTVIAVLEMFGNTDLAQALHEAISGRKQ"
     gene       complement(7652. .9712)
                /gene="katB"
     CDS        complement(7652. .9712)
                /gene="katB"
                /EC_number="1.11.1.6"
                /note="alternate gene name: katE"
                /codon_start=1

Query Match          3.2%;  Score 42.4;  DB 1;  Length 215534;
Best Local Similarity  64.0%;  Pred. No. 10;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy     77 ttaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgcccctcgatgaaaatca 136
          || |  |  ||| | ||  |   |    || || ||||| ||     ||  |||| | |
Db 122719 TTGCGCACATCGGGCTGGATACGGTGACATTTGACGGCACGCCTTTTTCCTTGAAGGTAA 122660

Qy    137 aggaaggtgaagccgtcaaaaaaggccaagtgctgtttga 176
          | ||||| ||  |||||||||||||| |||||||  ||||
Db 122659 AAGAAGGCGATACCGTCAAAAAAGGCGAAGTGCTTGTTGA 122620


RESULT  23
AF429315/c
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN


  Query Match          3.0%;  Score 40.2;  DB 9;  Length 125020;
  Best Local Similarity  10.6%;  Pred. No. 35;
  Matches   82;  Conservative  315;  Mismatches  367;  Indels    6;  Gaps    1;

Qy    303 gttcgaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgcaa 362
```

```
Db  17821 RKWWRWAYSMMRCAGMMGAACGASRSMKCCWGGYRKSAGSKSSRGGTGYYMKKKGGGKGS 17762

Qy    363 cctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccctgc 422

Db  17761 MSSKKWKGSSTSRRRGSSAKSCSYMWGMSMCMSCMSMAKSYMMCYCYMYRMSSYMSYYYK 17702

Qy    423 cgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggctgc 482

Db  17701 CYSCMGMSSTSYSYSCCWKMSSWGSCYKCMKYYSGWSSYSTSMGYYSSTSCKKYKCSWSS 17642

Qy    483 cgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggtatt 542

Db  17641 MYKCKTSKSYRRKRSYYYWGGGKRAKKKYYCAGRRRRMSYWKCCAKWWMSYCCWSYCMTY 17582

Qy    543 gagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgcagacgtgccgtc 602

Db  17581 YYSKSCTYKSSTCYKRGGYYWGSKTCYSAGGKSRSMYYCMMRSSKSSSWSMSMARSSWCM 17522

Qy    603 tgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcctgccggcttgag 662

Db  17521 GWGAGYRRSKRSAGWGAGWRSSKGKRSTGMKRACSKKTGSYGSTGRSMKKKKGYSKYSRG 17462

Qy    663 tggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtgtggaccatcaa 722

Db  17461 MKGKKKTCYCMWKYYKYRKTSMCWWYYMKSWGYKRYKRCCMKKKGCTGYRGMSSKKSGYK 17402

Qy    723 ttatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctgaataccgagcg 782

Db  17401 SMRGMRGSSYSTSCWKSCWGYSWMKCMKYMSYKKKRRSMRGSMSSKGWRGYAGRGCYSSS 17342

Qy    783 cgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgtaccgttttggg 842

Db  17341 MWSTRKRRSKCYSYKSYKKGRGKMKGWGGMKRGSKYWSSMKKMKRSSWSKCYSTKSYKSG 17282

Qy    843 tgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaaccgcgtgatttc 902

Db  17281 RRKSKGWGRSTKSKAKSSMRMAGSKCTYGSSYWSNNRRNNRMGKTGCNYMYRRSRAMMNG 17222

Qy    903 cggttcggtattgaacggtgcga------ttgcacaaggcgcgcatgattatttgggacg 956

Db  17221 NAAAAGCTTCCCCANTNGGGGGAAAAAGGGCGSASRASCYKGRMSSKSCYRSGTRRRCMK 17162

Qy    957 ctaccacaatcagatttccgttatcgaagaaggccgcagcaaagagctgttcggctgggt 1016

Db  17161 SKCRRGSRGKSMGMTRGRSGGKTSYSAKGSGRGCYYCWGWGRKGRKCMSSRKMKMYKSYY 17102

Qy   1017 tgcgccgcagccggacaaatactccatcacgcgcaccactctcggccatt 1066

Db  17101 RRKRWMTCMKMCYSMYMAMYCRSMCCMCMCKSCCGCYSMGMSSYSYSGKY 17052
```

```
RESULT  24
BHT1UL/c
LOCUS       BHT1UL                  37000 bp    DNA     linear   VRL 14-AUG-1996
DEFINITION  Bovine herpesvirus type 1 UL22-35 genes.
ACCESSION   Z78205
VERSION     Z78205.1  GI:1491620
KEYWORDS    capsid protein; DNA binding protein; DNA polymerase; envelope
            protein; glycoprotein B; glycoprotein H; proteinase; scaffold
            protein; tegument protein; thymidine kinase; UL22; UL23; UL24;
            UL25; UL26; UL26.5; UL27; UL28; UL29; UL30; UL31; UL32; UL33; UL34;
            UL35; UL36.
SOURCE      Bovine herpesvirus 1.
  ORGANISM  Bovine herpesvirus 1
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 37000)
  AUTHORS   Schwyzer,M., Vlcek,C., Lowery,D.E., Bello,L.J., Meyer,G., Misra,V.,
            Thiry,E. and Paces,V.
  TITLE     Gene contents in a 37-kb segment centered in the UL part of the
            bovine herpesvirus 1 genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 24211 to 27636)
  AUTHORS   Whitbeck,J.C., Bello,L.J. and Lawrence,W.C.
  TITLE     Comparison of the bovine herpesvirus 1 gI gene and the herpes
            simplex virus type 1 gB gene
  JOURNAL   Journal of virology. 62 (9), 3319-3327 (1988)
  MEDLINE   88300884
   PUBMED   2841484
REFERENCE   3  (bases 24215 to 27793)
  AUTHORS   Misra,V., Nelson,R. and Smith,M.
  TITLE     Sequence of a bovine herpesvirus type-1 glycoprotein gene that is
            homologous to the herpes simplex gene for the glycoprotein gB
  JOURNAL   Virology. 166 (2), 542-549 (1988)
  MEDLINE   89020821
   PUBMED   2845660
REFERENCE   4  (bases 27788 to 29695)
  AUTHORS   Haanes,E.J., Thomsen,D.R., Martin,S., Homa,F.L. and Lowery,D.E.
  TITLE     The bovine herpesvirus 1 maturational proteinase and scaffold
            proteins can substitute for the homologous herpes simplex virus
            type 1 proteins in the formation of hybrid type B capsids
  JOURNAL   Journal of virology. 69 (11), 7375-7379 (1995)
  MEDLINE   96013856
   PUBMED   7474173
REFERENCE   5  (bases 31487 to 32728)
  AUTHORS   Whitbeck,J.C., Lawrence,W.C. and Bello,L.J.
  TITLE     Characterization of the bovine herpesvirus 1 homolog of the herpes
            simplex virus 1 UL24 open reading frame
  JOURNAL   Virology. 200 (1), 263-270 (1994)
  MEDLINE   94174724
   PUBMED   8128626
REFERENCE   6  (bases 32460 to 33539)
  AUTHORS   Bello,L.J., Whitbeck,J.C. and Lawrence,W.C.
  TITLE     Sequence and transcript analysis of the bovine herpesvirus 1
            thymidine kinase locus
  JOURNAL   Virology. 189 (2), 407-414 (1992)
  MEDLINE   92351539
   PUBMED   1322582
REFERENCE   7  (bases 33521 to 36452)
  AUTHORS   Meyer,A.L., Petrovskis,E.A., Duffus,W.P., Thomsen,D.R. and
            Post,L.E.
  TITLE     Cloning and sequence of an infectious bovine rhinotracheitis virus
            (BHV-1) gene homologous to glycoprotein H of herpes simplex virus
  JOURNAL   Biochimica et biophysica acta. 1090 (2), 267-269 (1991)
  MEDLINE   92031707
   PUBMED   1657187
REFERENCE   8  (bases 36100 to 37000)
  AUTHORS   Vlcek,C., Benes,V., Lu,Z., Kutish,G.F., Paces,V., Rock,D.,
            Letchworth,G.J. and Schwyzer,M.
  TITLE     Nucleotide sequence analysis of a 30-kb region of the bovine
            herpesvirus 1 genome which exhibits a colinear gene arrangement
            with the UL21 to UL4 genes of herpes simplex virus
  JOURNAL   Virology. 210 (1), 100-108 (1995)
  MEDLINE   95313343
   PUBMED   7793062
REFERENCE   9  (bases 1 to 644)
  AUTHORS   Schwyzer,M., Styger,D., Vogt,B., Lowery,D.E., Simard,C.,
            LaBoissiere,S., Misra,V., Vlcek,C. and Paces,V.
  TITLE     Gene contents in a 31-kb segment at the left genome end of bovine
            herpesvirus-1
  JOURNAL   Vet. Microbiol. (1996) In press
REFERENCE   10 (bases 1 to 37000)
  AUTHORS   Schwyzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1996) Martin Schwyzer, Institute of Virology,
            Faculty of Veterinary, Medicine, University of Zurich,
            Winterthurerstr. 266a, Zurich, CH-8057, Switzerland
FEATURES             Location/Qualifiers
     source          1. .37000
                     /organism="Bovine herpesvirus 1"
                     /strain="Cooper"
                     /db_xref="taxon:10320"
                     /clone="HindIII fragments E (partial), G, and A (partial)"
     misc_feature    2. .7
                     /note="PstI site located at base 30802. .30807 from left
```

```
                genome end"
CDS             108. .9851
                /note="very large tegument protein"
                /codon_start=1
                /product="UL36"
                /protein_id="CAB01605.1"
                /db_xref="GI:1491621"
                /db_xref="SPTREMBL:Q65553"
                /translation="MAFAPVARPQPGEARAGGAAGGLRINGAARAVPPSTPVEQRSGT
                MSGVAGARAPATPADVAVAAVGFRNQYDAALGPGSAVACLRSSLSFLRLAFAGGVDAA
                LGADAIDGALAEGAAWARESGARPEMCSIVHLPNRITDQAGAGSGLCCVFSRVYGECG
                FYTPPAERALSTQVPAREFVDAVWQPRRTSLALVIVGAMGVGVFRAGDAVYLFDPHGS
                GDVPQAFVARMPPRALYAYLALRACGQPESRWAGALVHFVSAGPEPARPEELRAAVSA
                LYGASETYLDDEPYVERRVVSVHPRAPANTACLTAVAVGGAGAECEDPTFARPGSPAL
                LPAEPVARMPPQITAAAAARAAGKRTSLPRRRRAPWTPPSSREDLRAPAARARAGKPP
                RKVRGAETAGAGEDWGEGGNAAAARPPPALGAAAPAHAATACAAELGARAAEVTAALD
                AVQAHAASIAADADACVARAVRVARPRDAAAERDPLEHAIAGLFERVLAFLVQNGART
                RHDAPSAVQPLFGAVAQALPQHTPASAFITSSGMVLEGLAAYTRLSEAMRAQPSALGD
                LADAKLRLVAAAVAADTERLHDAVDELEGEVDGAGGAGPLELCGAVSARFMERLSASG
                RRLFSRAALRDGRPLDARVRALFERARAREARVAAETAALAREILALEGTVRAARERL
                DAVPPDDPRAVPADSAVADLAAELAAGRAAVVARSAAALSAELDALAAEALRQYFLRG
                AQYSARAILADRGGGERFRVASAAVAPLQRLAGSAPDLAARVARLGAAPGAAAPPLPR
                SARKGELLRDLLAAGATAHTEDGLAAWVALLHEAQAEGAAAKKELDAVLSDVAAINGR
                SASRASLEASMARFEAMSAAAARAAAGAAAEAPGAGAEEAADAAEAAVRGADDVLRQL
                DALATGAAADREARARVAAARAGLEAQALAARRRVAELGERRAELYQRLDALLLPLQG
                FAGLRAAPGALERLCADAGAQSAEDCARFLQEAPPRVAAGVQDRLWQLFGRYREALEN
                PSALAPGALAGLGPAFAAVLGTALGQAMGPAARAFFEGHAERVAAAVARAAAEPAAPA
                AVAAAASALREAAEALRAGGAHVPPEFAFLDALQAQYAARREAQTGARRLGAALAAVS
                EASGALASAASALKNAMTAGADAAETTAAAAAADAALGAAEEVLRAADEAAVAAGGED
                ADGGEDASEADADGAGDAPPAGRLGPQAAKLHADLRKARALARRRAEEVRRLRAEAER
                RAEGAAAQRQEERWREDLRAALERVEANSAFDAAELARLRAAAAARGYDPRPLARQAD
                RALEASAQRATAAIEAVLAFNPHAPENAKIDVPPPLAQLRGIAWWDAFSLAAPALSAL
                FPAADVGQLTRLMHIATGLLTFAASGGGEPRYYDAVTYLEGDLAAVPRLAKYVAFYRR
                GHAAFEEERARLGALRADVLQAAGGRAAEISRARRGNVLRSPAERRALEAGVRLELPS
                EALIASALAQLERFDHARFAGSAYEAQMLSTVRRDAAAAREALEAAKAARAEATARAE
                RILGEVVAAEAARDRDEGAGLANLKNLLRITPPPAALGPALDRAASAADVVTQAALLL
                AAVEEAPELDVAAVEWLQQARSIIDSHPLTARIDDRGPMEPFAERIDALRELRRHLDA
                LRRQLESAAAAWDGAWELFVAARRRAAASREDHEGAKARAAALQAAAGVVLGLRADEH
                YARLPAAFTGALDARLAERGDALGAFHEAARAADAGLQRLQATLAAVRGEAAYEGLRA
                LLAAFDAQVAELPAWAAQAHAPFRDLLVLRMRLYEAYFKLVPRAAARARGAAAARAAR
                AAATVHSRADPATRAVAARAARLCERDLHLRGRVAALLGDEAAVCTLREADSEIDALA
                PRAYLDADGTPLCYRVCYRAVGDKLAVALCSQVGASMRPPLDESGVVEAASVAAVNVL
                NEIVNLRLEIERARDADFGAFCRFVRHRRADWGLADVRSAAAELYAGLLATTLPRRHG
                VASLGALCFSLAERGLRPARAPASGPAPSGGGALELTPADLVVAALMGAFMHLVNFTR
                LDLVQKQEYMCKTLDGVLTEALTSRVAVNSLAPGEPGARRPVPLGGGADDPADGALFA
                IRAADWDAGRLSPSDVLAVWRHAPDAATRATVAELARLIPGGALTTISVLARMCIPGD
                LLAALWTTLAVDSLGAQTQSYDAFLARRLDAPSTVHATGGPSEDAAALAAAAAAAGGR
                PALFCPTGSSVTFTLVRQPPSEVQTVNALDLVTCALLLGAPVVVAMENPDVFSEGSRL
                IMCLRLFDTRPGGRDADAPAAVSSDLNSWGERLLALDENIIENACLTAQLEQLSALIA
                SKPLRSAPPCLIMLDTHLQLARVLWARAAAPPLAMLKIAEDDVLRDLPSLTLDYEDAL
                PPAASAADPLFTHIISENNVPDFAAASGDPLYAHPPVFQHATDNLFPHARAAAGGGSG
                NRVANARAVAAAARPPTPASAPAPAPALAPASAEPALAWGADEWLDDEIGDAYPPAPA
                HDAYPPAPAHDAYPPAPAHAQGKPALLAAPAAGAAEAPGVERARRLTRRTGPRKSMPA
                ALPWRRPPPASLVSAAPPVPPGPRLPPAPPLPPPAPPLPPPAPPLPPPAPPLPPPAPP
                LPPPAPPLPPPAPSTAPVPAPPLPPPALTPALTPAPTPAPTPAPPLPLPAPITVLVPA
                PVPAPAPIPAPAPTPAPAPTPAPPLPPPAPDGAMGALSATRRPTRRAGARKSLPAAQP
                RQRLLRSRSPASVPAPGSELVPPPSGGALGSPPSFVPSRPPSLEPVPGLPLPPSRVQA
                PVDAPAPPPAPERPAPPPAPERPAPPPAPERPAPPPAPERPAPPPAPERPAPPPAPER
                PATAVQSGSKIPAPSNALVGQVAPALLRRGPSPRGKPGAPARPTGGANIARKNTRAVD
                PTVIGSGSALSGLGSAEEDADFGASGMYVAPPGPAPLAAPPAPSALPAPRLEAPAAAS
                LPAPPIAPPIAPPIAPPIAPPIAPPIAPPIAPPIAPPIAPPAMPAPPSAAPATAGSSA
                PPIAPPATPAPLIAPLAAPVPRLVGAAARGRVIGHMRAAAPKPRGLRAAERPPMRSWA
                SRGNLHTQDLLASAEARICELPAPPKVPLGADSDATDSSLDGSGPETEEETEAEAGGD
                GPGQDAPKARVVPANSLLKRQYLRGTGISVLALLLEACEKIARRLRATRDVLRQRAAE
                VTADIFALRLLLG"
misc_feature    639. .644
                /note="EcoRI site in HindIII E fragment"
repeat_region   7697. .7777
                /rpt_type=DIRECT
                /rpt_unit=7697. .7723
repeat_region   8609. .8770
                /rpt_type=DIRECT
                /rpt_unit=8609. .8635
misc_feature    9110. .9115
                /note="HindIII E/G boundary"
repeat_region   9120. .9227
                /rpt_type=DIRECT
                /rpt_unit=9120. .9131
polyA_signal    9928. .9933
                /note="UL37 and UL36"
polyA_signal    complement(9969. .9974)
                /note="UL35, UL34, UL33"
CDS             complement(10012. .10386)
                /function="capsid protein"
                /codon_start=1
                /product="UL35"
                /protein_id="CAB01606.1"
                /db_xref="GI:1491622"
                /db_xref="SPTREMBL:Q65554"
                /translation="MSAPAGIDPSAPATITPDTLRDLLLPVQILHVLNAAARPLQDGTT
                PEQVSAARRNLLVGTSLAMVDLRRRHEKAVVPRVPMFATYDHAHWARPTIGLKRTFLP
                RVVQLPPEYEEDAGAAASAART"
CDS             complement(10431. .11213)
                /note="envelope protein; putative"
                /codon_start=1
                /product="UL34"
                /protein_id="CAB01607.1"
                /db_xref="GI:1491623"
                /db_xref="SPTREMBL:Q65555"
                /translation="MHGRERDCDGEGEGEGSSRSLMARIRLLVAGNLQCGEGDVPQPW
                DPRRPPARCVFQFNGQDGSNESFPLEYILRLMASWAQTDCDPYVRVQNTGVSVLFQGY
                FSRPPGAPLAAITAEQNNATLASTQSTGLSLSALEKIKARGGIDPRPFRAMMSVSCFV
                RMPRVQLSFRFMGPGDSHRTGRLLDRAVRERLSERDVRSRRASACGHSERELEPAGGA
                GNVAHYAGALLTARRQKMAALATLVGGVCGILAILRLFWAPS"
TATA_signal     complement(11285. .11291)
                /note="UL34"
CDS             complement(11324. .11650)
                /note="involved in genome cleavage/packaging"
```

Query Match 3.0%; Score 40; DB 14; Length 37000;
Best Local Similarity 51.7%; Pred. No. 36;
Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy    766 cgtctgaataccgagcgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctc 825
          || |||| | |||||||||    | ||   | ||||| || |   |  | |   |||| |
Db  31007 CGGCTGATTGCCGAGCGCCCTCTGGCGACGCGCGGCATGGACTACCGCGATGGGCGCATG 30948

Qy    826 ttgcgtaccgttttgggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcg 885
          |  || ||| |  |||  ||   |||| |    ||     || |||      | ||| |
Db  30947 TCTCGAACCTTCATGGCGGCCGCGGTGGCGGCGCTGCAGTCCAGCGGGCGCATGTACGTG 30888

Qy    886 gacaaccgcgtgatttccggttcggtattgaacggtgcgattgcacaaggcgcgca 941
          | ||||||||  |   || |  |  |  |     |    |  |   ||   ||||||
Db  30887 GGCAACCGCGCGTACTCGGCCTTTGAGTGCGCGGTGCTGTGTCTGCACCTCGCGCA 30832
```

```
RESULT 25
BHV1CGEN/c
LOCUS       BHV1CGEN              135301 bp    DNA     linear   VRL 07-APR-2000
DEFINITION  Bovine herpesvirus type 1.1 complete genome.
ACCESSION   AJ004801
VERSION     AJ004801.1  GI:2653291
KEYWORDS    complete genome.
SOURCE      Bovine herpesvirus type 1.1.
  ORGANISM  Bovine herpesvirus type 1.1
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 99702 to 103033)
  AUTHORS   Wirth,U.V., Fraefel,C., Vogt,B., Vlcek,C., Paces,V. and Schwyzer,M.
  TITLE     Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
            are 3' coterminal and encode a putative zinc finger transactivator
            protein
  JOURNAL   J. Virol. 66 (5), 2763-2772 (1992)
  MEDLINE   92219360
REFERENCE   2  (bases 103034 to 111027; 127191 to 135300)
  AUTHORS   Schwyzer,M., Vlcek,C., Menekse,O., Fraefel,C. and Paces,V.
```

```
  TITLE     Promoter, spliced leader, and coding sequence for BICP4, the
            largest of the immediate-early proteins of bovine herpesvirus 1
  JOURNAL   Virology 197 (1), 349-357 (1993)
  MEDLINE   94025583
REFERENCE   3  (bases 111028 to 114234; 123984 to 127190)
  AUTHORS   Schwyzer,M., Wirth,U.V., Vogt,B. and Fraefel,C.
  TITLE     BICP22 of bovine herpesvirus 1 is encoded by a spliced 1.7 kb RNA
            which exhibits immediate early and late transcription kinetics
  JOURNAL   J. Gen. Virol. 75 (Pt 7), 1703-1711 (1994)
  MEDLINE   94292919
REFERENCE   4  (bases 66900 to 96900)
  AUTHORS   Vlcek,C., Benes,V., Lu,Z., Kutish,G.F., Paces,V., Rock,D.,
            Letchworth,G.J. and Schwyzer,M.
  TITLE     Nucleotide sequence analysis of a 30-kb region of the bovine
            herpesvirus 1 genome which exhibits a colinear gene arrangement
            with the UL21 to UL4 genes of herpes simplex virus
  JOURNAL   Virology 210 (1), 100-108 (1995)
  MEDLINE   95313343
REFERENCE   5  (bases 1 to 31444)
  AUTHORS   Schwyzer,M., Styger,D., Vogt,B., Lowery,D.E., Simard,C.,
            LaBoissiere,S., Misra,V., Vlcek,C. and Paces,V.
  TITLE     Gene contents in a 31-kb segment at the left genome end of bovine
            herpesvirus-1
  JOURNAL   Vet. Microbiol. 53 (1-2), 67-77 (1996)
  MEDLINE   97164286
  REMARK    Accession# Z54206
REFERENCE   6  (bases 30801 to 67800)
  AUTHORS   Schwyzer,M., Vlcek,C., Lowery,D.E., Bello,L.J., Meyer,G. and
            Misra,V.
  TITLE     Gene contents in a 37-kb segment centered in the UL part of the
            bovine herpesvirus 1 genome: the last gap
  JOURNAL   Unpublished
  REMARK    Accession# Z78205
REFERENCE   7  (bases 96901 to 99695)
  AUTHORS   Letchworth,G.J. and Kutish,G.F.
  TITLE     DNA sequence of the BHV-1 UL1 to UL3.5 genes
  JOURNAL   Unpublished
REFERENCE   8  (bases 114235 to 122983)
  AUTHORS   Goltz,M., Buhk,H.J., Broll,H., Lewin,M., Mankertz,A., Boerner,B.,
            Borchers,K. and Weigelt,W.
  TITLE     Nucleotide sequence of the HindIII O and K fragments located in the
            US region of the bovine herpesvirus 1 genome
  JOURNAL   Unpublished
  REMARK    Accession# Z98199
REFERENCE   9  (bases 121402 to 123983)
  AUTHORS   Schwyzer,M.
  TITLE     Glycoprotein E and US9 genes of BHV1
  JOURNAL   Unpublished
REFERENCE   10 (bases 1 to 135301)
  AUTHORS   Schwyzer,M., Paces,V., Letchworth,G.J., Misra,V., Buhk,H.J.,
            Lowery,D.E., Simard,C., Bello,L.J., Thiry,E. and Vlcek,C.
  TITLE     Complete DNA sequence of bovine herpesvirus 1
  JOURNAL   Unpublished
REFERENCE   11 (bases 1 to 135301)
  AUTHORS   Schwyzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1997) Institute of Virology, Faculty of
            Veterinary Medicine, University of Zurich, Winterthurerstrasse
            266A, Zurich CH-8057, Switzerland
FEATURES             Location/Qualifiers
     source          1. .135301
                     /organism="Bovine herpesvirus type 1.1"
                     /db_xref="taxon:79889"
                     /focus
     intron          <1. .472
                     /note="3'part of intron spliced out from transcript over
                     covalently joined genome ends, leading to immediate-early
                     RNA 1.5 and circ protein"
     source          order(1. .4014,102918. .111027,127191. .135300)
                     /organism="Bovine herpesvirus type 1.2 (strain K22)"
                     /virion
                     /strain="K22"
                     /db_xref="taxon:31519"
     source          1. .135301
                     /organism="Bovine herpesvirus type 1.2"
                     /db_xref="taxon:79890"
     repeat_region   235. .430
                     /rpt_type=DIRECT
                     /rpt_unit=235. .248
     mRNA            466. .1280
                     /note="transcribed from alternative late promoter"
                     /product="LR1.1 (circ)"
                     /evidence=experimental
     CDS             486. .1229
                     /note="immediate-early and late myristylated protein
                     (virion)"
                     /codon_start=1
                     /product="circ protein"
                     /protein_id="CAA06078.1"
                     /db_xref="GI:2653292"
                     /db_xref="SPTREMBL:Q01342"
                     /translation="MGARASAPAAGPPPAHAVLLDALSGGTIDLPGGDEAVFVSCPTT
                     RPVYHHMRRGRTAHTTPVHFVGRAYAILPCRKFMLYLMRGGAVYGYEPTTGLHRLADS
                     LHDFLTTAGLQQRDLHCLDVTVLDAQMDPVTFTTPEILIELEADPAFPPPPSARARRS
                     TLRRASMRRPARTFCPHQLLAEGSILDLCSPEQAAAPGCSLLPACDSGDAACPCDAGE
                     TARDCTADAARAPSPGALSRYSSVRSVFF"
     polyA_signal    1259. .1264
                     /gene="circ"
     gene            1259. .1264
                     /gene="circ"
     gene            complement(1649. .1654)
                     /gene="UL52"
                     /note="UL53; UL54"
     polyA_signal    complement(1649. .1654)
                     /gene="UL54"
     gene            complement(1658. .2860)
                     /gene="UL54"
     CDS             complement(1658. .2860)
                     /gene="UL54"
                     /note="early BICP27 affecting mRNA processing (cell
                     nucleus)"
                     /codon_start=1
                     /product="UL54 protein"
                     /protein_id="CAA06079.1"
                     /db_xref="GI:2653293"
                     /db_xref="SPTREMBL:Q01354"
                     /translation="MADPEIATLSTASESDDLSLFGSDREEDDEAPSLAPALRSVVGQ
                     VRKRKLEGAEDEPMPAEPPGEGAASGDGGPAEAPPARRARVRPRRPRRRPRRRQPAGE
                     QRSRGPAAKREAALATSSHGGGGAAARSIGSSLRLARSLAEAAQRATAERVTAVFAGA
                     RLDLMRPVQNGGFRAAGVSPWAAVLDFGAEQFVPEGRRVTWETLMFHGRDLYRMFEVR
                     PHAAQAARALRDLVLRSANLVDALASADECLTWCKFIATKNLRLRTKDPIVATAGAVL
                     ENLRLKLAPFLRCYLRGRGLPSLEELCAARRLSLATCPASYMFVMLARLSRAVRSGAE
                     CVPLLEVTVGDAPFEEYIPGTCVAGLIDALDTHKQACDSMTCKLVANFTLVPVYMHGK
                     YFYCNEIF"
     gene            complement(3040. .4038)
                     /gene="UL53"
     CDS             complement(3040. .4038)
                     /gene="UL53"
                     /codon_start=1
                     /product="glycoprotein K"
                     /protein_id="CAA06080.1"
                     /db_xref="GI:2653294"
                     /db_xref="SPTREMBL:Q65816"
                     /translation="MLLGGRTVNLAALALLTAHLALALWVALAARCQRCACVRATARN
                     GSLRWELRSPGAVYVWGGANNATLAADAPCRHAVVQHIPPGLLDGDEALHGRVRAVAG
                     ARDCRAYLWCAQARGGLLAWLLYVAFVYLRQERRMFGLCRNDADFLSPGGYTLNYAAA
                     ALAAVVGHGPYTKLARLMCELSARRRALAVDFRLDPLGCAWRPRAALPLLAEGFARLG
                     ARIAAAGSVGITHPCAAAYPLYLKIWAWVHVALFAGLELVSLLYRKPRRRGGTCAGDG
                     GDGGESGIRKVCVNCCSTLLAGLLVKALYLAAIVGGVIALLHYEHNLRLRLLGAQT"
     gene            complement(4013. .7237)
                     /gene="UL52"
     CDS             complement(4013. .7237)
                     /gene="UL52"
                     /codon_start=1
                     /product="component of DNA helicase/primase complex"
                     /protein_id="CAA06081.1"
```

```
                     /db_xref="GI:2653295"
                     /db_xref="SPTREMBL:Q65817"
                     /translation="MDGPAALEDALQILYATDGCAIGFSLMLLTGQEEPGRGAYVVSY
                     NWPGRLAACLGGESPNDVALALADARPLVAFYLLGGPPGGSAAATPRLARLRVHLQQR
                     AGRARARRRTAPRQGPTRARAARAPRRGRHLCAARGHDHALLVATEQLAPRTGRTADD
                     ARYVEGQTTVRSALRARPAGRRGLASLYIHHEHKTVAAYRRLYSNSGATPFWFLSKFG
                     PGEKTLVLATRFYVFQAERAGDAATYDLQAVRDCLATYAVAAPPNPSGLAFPDLVSFA
                     ALAAFCCRSGYARGAVAAGAPAYVAARIEADLAEVRCLREYIDHDRRSLKVADPEFVA
                     YVYLAYFEGFNRRQITEHLRAVTATEPPDGGEALAPRLAGVSRLRERAVDAFFRHVRA
                     QFNVQSYIEQNVAVAVVRLPPATAEAYARARTYARLLAATANGPGRTICDGAAALRAA
                     LDHLEGQAARFGWVLHAPGAPSASAVPGPGLDAALRVGAPDGGASPDATPACCGVSKR
                     LLELAAAAARDGAHPLDGLFGARGTAAPTPVYRVELPRGHQAFAVADGDDWAAVTSVA
                     ALDPDEVAVEVAAAAAAAAESGRDALVERDARLTALLTERAACAAGRSDRRSLGPAAP
                     RDQYYVNRNELFNARLAVTNIVLDVDFRLKRPLPRGDLHGAMRSFRRGALAALAALFP
                     EIAADAWPAHPCFFYKSACPSAAASNLGDGDASPSWAAPDRDAGWDEYESALAREGEA
                     QYSCEDVSDPAGARPEETGGAEGAGSAAALTTACGCDDKMGFRVAVPVPAPYIIAGGP
                     TLKGIARLVQHAVMLERPFAEAMSRYLRDFAFVDAAVYTHGHSLRLPFFGKPDAASNV
                     GRWLLPFYVVPERCDDVAAFVAAHRDPHHFHFHAAPPAGGRREARRAHVGGEYVSFFE
                     RKAAVNRAAFHGVRVSLRRALAAVGVDGGDRGAVEEFIADVIEAELAPYMAEHYPQAA
                     REYQGAGARLVAAKADWLLAQLVPARAGTAQGFGCLRAAHGRAAQNKARSLVSLSVDS
                     HDRLCASLIQQCFATKCGSNRLGTVFTVDLSGRAARYAARGADC"
     source          order(4015. .9664,15010. .22094,26636. .102917,
                     114235. .123983,135301)
                     /organism="Bovine herpesvirus type 1.1 (strain Cooper)"
                     /virion
                     /strain="Cooper"
                     /db_xref="taxon:10323"
     gene            join(7236. .7967,7976. .7981)
                     /gene="UL51"
     CDS             7236. .7967
                     /gene="UL51"
                     /codon_start=1
                     /product="virion protein"
                     /protein_id="CAA06082.1"
                     /db_xref="GI:2653296"
                     /db_xref="SPTREMBL:Q65818"

  Query Match              3.0%;  Score 40;  DB 14;  Length 135301;
  Best Local Similarity   51.7%;  Pred. No. 40;
  Matches   91;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     766 cgtctgaataccgagcgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctc 825
           || |||| | |||||||||    | ||   | ||||| || |   |  | |   |||| |
Db   61807 CGGCTGATTGCCGAGCGCCCTCTGGCGACGCGCGGCATGGACTACCGCGATGGGCGCATG 61748

Qy     826 ttgcgtaccgttttgggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcg 885
           |  || ||| |  |||  ||   |||| |    ||     || |||  .   |  ||| |
Db   61747 TCTCGAACCTTCATGGCGGCCGCGGTGGCGGCGCTGCAGTCCAGCGGGCGCATGTACGTG 61688

Qy     886 gacaaccgcgtgatttccggttcggtattgaacggtgcgattgcacaaggcgcgca 941
           | |||||||| |   || |  |  |  |     |   |  |   ||   ||||||
Db   61687 GGCAACCGCGCGTACTCGGCCTTTGAGTGCGCGGTGCTGTGTCTGCACCTCGCGCA 61632


RESULT  26
AL646076
LOCUS       AL646076              216050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
            segment 1/11.
ACCESSION   AL646076 AL646053
VERSION     AL646076.1  GI:17430467
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 216050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .216050
                     /organism="Ralstonia solanacearum"
                     /plasmid="megaplasmid"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     rep_origin      1. .17
                     /note="ori or RS05845
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   52. .68
                     /note="RS05846
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(77. .93)
                     /note="RS05847
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   159. .175
                     /note="RS05848
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   202. .218
                     /note="RS05849
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(227. .243)
                     /note="RS05850
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(263. .279)
                     /note="RS05851
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(307. .323)
                     /note="RS05852
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(374. .390)
                     /note="RS05853
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(516. .532)
                     /note="RS05854
                     HYPOTHETICAL PARA BOX
                     predicted by Homology"
                     /evidence=not_experimental
```

```
repeat_region   complement(670. .686)
                /note="RS05855
                HYPOTHETICAL PARA BOX
                predicted by Homology"
                /evidence=not_experimental
repeat_region   complement(733. .749)
                /note="RS05856
                HYPOTHETICAL PARA BOX
                predicted by Homology"
                /evidence=not_experimental
repeat_region   complement(853. .869)
                /note="RS05857
                HYPOTHETICAL PARA BOX
                predicted by Homology"
                /evidence=not_experimental
repeat_region   complement(919. .935)
                /note="RS05858
                HYPOTHETICAL PARA BOX
                predicted by Homology"
                /evidence=not_experimental
gene            1219. .2244
                /gene="repA"
                /note="RSp0001; RS02238"
CDS             1219. .2244
                /gene="repA"
                /function="elements of external origin; plasmid-related
                functions non-bacterial functions"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE REPA REPLICASE PROTEIN"
                /protein_id="CAD17152.1"
                /db_xref="GI:17430468"
                /translation="MPGRIIVPSGGQVIATPEKFQRLFDEAQAMEREDAWQSGEVGFL
                SRASVQVTLPYRAPKGSPPVWTRSSGNISLMIQPGYFTQQRSERASNGRQRIVSETVS
                FGYPYGSYPRLMLAWIGKEIMAKKKRGEFQGTVEDRRISLGNSLSEFMYNLGIPMATG
                GKRGTMTLVRQQMIRLFSATIAIVQNKSAPNNQNQEPLSIDRIGYLLADQLSTWWDPM
                QPGQGSMFESFVVLSEPFFNELVNRPVPVDMRALKALKQSPFALDVYSWLTYRFFTIQ
                KRTEIPWEALQMQFGTETESERKFRALFRKALKDVLVVYPDAKVDADSSKALILQPSR
                TSVRKLA"
gene            complement(3120. .3323)
                /gene="cspD3"
                /note="RSp0002; RS05790"
CDS             complement(3120. .3323)
                /gene="cspD3"
                /function="miscellaneous; not classified regulator"
                /note="Product confidence : probable
                Gene name confidence : putative
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE COLD SHOCK-LIKE TRANSCRIPTION REGULATOR
                PROTEIN"
                /protein_id="CAD17153.1"
                /db_xref="GI:17430469"
                /translation="MATGTVKWFNETKGFGFITPDGGGADLFAHFSEIQGSGFKTLKD
                GQKVSFEVKQGPKGLQASGIKPE"
gene            3728. .4396
                /gene="parA"
                /note="RSp0003; RS01975"
CDS             3728. .4396
                /gene="parA"
                /function="elements of external origin; plasmid-related
                functions non-bacterial functions"
                /note="Product confidence : probable
                Gene name confidence : putative
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE PARTITION PROTEIN"
                /protein_id="CAD17154.1"
                /db_xref="GI:17430470"
                /translation="MSAKIVTVFNQKGGCGKTTVSMHIAGTLGVRGSKTLLVDMDEQG
                TATRWAAQAPEEKPFPASVIGLAPSGGAMHREVRKFIADYDYIVVDCPPAVHSPASSS
                ALLISDIAIIPVVPSPPDLWAAVAAKALAQQAQVTNETLRVRVLANMVQRRVSLAKQA
                IEILGDDGDIPLMESMIGSRSAFRECQAIGATVHGVPGAREAINEVDIMVNEVLQLLR
                EGQA"
gene            4393. .5397
                /gene="parB"
                /note="RSp0004; RS01976"
CDS             4393. .5397
                /gene="parB"
                /function="macromolecule metabolism; macromolecule
                synthesis, modification; dna - replication, repair,
                restr./modif."
                /note="Product confidence : putative
                Gene name confidence : putative
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PUTATIVE PARTITIONING PROTEIN"
                /protein_id="CAD17155.1"
                /db_xref="GI:17430471"
                /translation="MSAKLKSLKDGMLAGIAAEKTKAAPVDRFALAEQAITSHPRGLL
                GQPARPAFSAESQAETSSERRVVKIPLERLQENPLNARRIYDQQIVQERAASIATHGQ
                QTPGLAAPDPAKPGWYVLIDGHYRKRALAAAGKAEMECFIEDGLSDIDFYRLSFVLNE
```

```
  Query Match             3.0%;  Score 39.8;  DB 1;  Length 216050;
  Best Local Similarity   48.5%;  Pred. No. 46;
  Matches  110;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    1085 tcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggcacttatg 1144
           || |   || | | | || || || || | ||| | ||   |         |||     |
Db   20800 TCGACCACAAGGCGGGCGACATCGCCGACAACCTCTCCTATGGCGACCAGCGCAAGCTCG 20859

Qy    1145 agcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtcggcg 1204
           ||| ||    | ||||| || |  ||| |   |||| ||| |||      ||||| ||
Db   20860 AGCTCGCGCGGGCGTTGGCCACCGAGCCGAAGCTGCTGCTGCTCGACGAGCCCGTCGCCG 20919

Qy    1205 ataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctcgctttgt 1264
             |  ||| ||   ||    | | |    |||    |  |||  |   ||  ||
Db   20920 GCATGAACACCGGTGAGAAGGTCGATCTGATGGCCGAGATCGAGAACATCCGGGCGCGCG 20979

Qy    1265 gcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgc 1311
           ||  |  | ||| |  |  | |  ||||     ||    || || ||
Db   20980 GCTACACCATCTTCATGATCGAGCACGACATGCGCTTCGTGATGGGC 21026


RESULT  27
AF429315
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens                                              .
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
```

```
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN


  Query Match             2.9%;  Score 39;  DB 9;  Length 125020;
  Best Local Similarity   10.3%;  Pred. No. 70;
  Matches   56;  Conservative  245;  Mismatches  240;  Indels    3;  Gaps    2;

Qy     352 gtgcgccgcaacctgattcaatcaggcttatggactgcgcttcgcacccgtccgttcagc 411
           ::: ||  |    :     :  |: |:   : : :  :::  :   ::: :|:   :|:
Db   17231 RKRNGCAMCKYNNYYNNSWRSSCRAGMSCTKYKSSMTMSMASYCWCMSMYYCSMRSMASR 17290

Qy     412 aaaatccctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaat 471
           : : :::: : : :::  :::   | : | |:: :  :: :  : :  ::: :: |  |
Db   17291 GMSWSSYMKMMKSSWRMSCYMKCCWCMKMCYCMMRSMRSRGMSYYMYASWKSSSRGCYCT 17350

Qy     472 ccgctggctgccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggc 531
            |    ::::|:: ::: : :: :  : ::: :| :  : : :    |::::: : :::
Db   17351 RCYWCMSSKSCYKSYYMMMRSKRMKGMKWSRCWGSMWGSASRSSCYKCYKSMRCSMMSSK 17410

Qy     532 ctgttggtattgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagcaggcgca 591
           |::  |  : :| |:::::  :::::     |:: ::  :: : |:: |::| ::::
Db   17411 CYRCAGCMMMKG-GYMRYMRCWSMKRRWWGKSAMYRMRRMWKGRGAMMMCMKCYSRMSRC 17469

Qy     592 gacgtgccgtctgaaaatgctgccaatatcgaaacacatgaatttggcggcccgcatcct 651
           :: :::|  | : |:: | :  || :: ::  :| |:      : :|   | |   : :
Db   17470 MMMMKSYCASCRSCAMMSGTYMKCASYMCMSSYWCTCWCTSYMSYYRCTCWCKGWSSYTK 17529

Qy     652 gccggcttgagtggcacgcacattcatttcatcgagccagtcggcgcgaataaaaccgtg 711
           : : ::  :  ::| : :: :  |::   : : :  :::| ::    ::  ::|   ::|
Db   17530 SKSWSSSMSSYKKGRRKSYSMCCTSRGAMSCWRRCCYMRGASSMRAGSMSRRRAKGRSWG 17589

Qy     712 tggaccatcaattatcaagacgtgattgctatcggacgtttgttcgtaacaggccgtctg 771
           ::   :|  ::  ::     | ::   :   | :: :: :   :: : ::::: : :

Db   17590 GRSKWWMTGGMWRSKYYYYCTGRRMMMTYMCCCWRRRSYMYYRSMSAMGMRKSSWSGMRM 17649

Qy     772 aataccgagcgcgtggttgccttgggcggcctgcaagtcaacaaaccgcgcctcttgcgt 831
           :  |::::  : : :::  |:   :|::|:|::: :    ::: |:::|: :    : ::
Db   17650 MGSASSRRCKSASRSSW--CSRRMKGMRGSCWSSKMWGGSRSRSASSKCKGSRGMRRRSK 17707

Qy     832 accgttttgggtgcgaaggtgtctcaacttaccgccggcgaattggttgacgcggacaac 891
           ::::: : |:|:: :: :::|      |:::: :: ::|     ::   :::::: :  |
Db   17708 RSSKYRKRGRGKKRSMTKSKGSKGKSKCWKRSGSMTSSCYYYSASSCMWMMSSKSCMCCC 17767

Qy     892 cgcg 895
           : ::
Db   17768 MMMK 17771


RESULT  28
AB047272
LOCUS       AB047272               8918 bp    DNA     linear   BCT 14-FEB-2001
DEFINITION  Pseudomonas sp. CA10 cat operon (catR, catB, catC, catA, ORFcat5,
            ORFcat6, ORFcat7, ORFcat8), complete cds.
ACCESSION   AB047272
VERSION     AB047272.1  GI:12862416
KEYWORDS    .
SOURCE      Pseudomonas sp. CA10 (strain:CA10) DNA.
  ORGANISM  Pseudomonas sp. CA10
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (sites)
  AUTHORS   Nojiri,H., Maeda,K., Sekiguchi,H., Nakai,S., Yoshida,T., Hiroshi,H.
            and Omori,T.
  TITLE     Analysis of carbazole degradation genes from Pseudomonas sp. strain
            CA10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 8918)
  AUTHORS   Nojiri,H., Omori,T. and Habe,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2000) Hideaki Nojiri, The University of Tokyo,
            Biotechnology Research Center; 1-1-1 Yayoi, Bunkyo-ku, Tokyo
            113-8657, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp,
            Tel:81-3-5841-3064, Fax:81-3-5841-8030)
FEATURES             Location/Qualifiers
     source          1. .8918
                     /organism="Pseudomonas sp. CA10"
                     /strain="CA10"
                     /db_xref="taxon:135214"
     gene            complement(594. .1475)
                     /gene="catR"
     CDS             complement(594. .1475)
                     /gene="catR"
                     /function="Positive regulator of cat operon"
                     /codon_start=1
                     /transl_table=11
                     /product="positive regulator of cat operon CatR"
                     /protein_id="BAB32455.1"
                     /db_xref="GI:12862417"
                     /translation="MELRHLRYFQALAETLNFTRAAERLHIAQPPLSRQIQQLEDELG
                     VALLERGRPLRLTEAGRFFHEHSSTLLEQLTKVCDNTRRIGSGEKAWLGIGFAPSTLY
                     GVLPDLIRRLRSDAGLELGLSEMTTLQQVESLKAGRIDVGFGRIHINDPAIQQQVLTL
                     DPLVAALPAGHPLLPGPVSLTDLAQESFVLYPGNPRPSYADHVLALFAAYGLQIEVAQ
                     WTNELQTAIGLVAAGIGITLVPASVQVLHRADIGYAKLQEAQATSPIIISRRTGDASP
                     ALTHCLALIEELSRGRL"
     gene            1612. .2733
                     /gene="catB"
     CDS             1612. .2733
                     /gene="catB"
                     /function="cis,cis-muconate cycloisomerase"
                     /codon_start=1
                     /transl_table=11
                     /product="cis,cis-muconate cycloisomerase CatB"
                     /protein_id="BAB32456.1"
                     /db_xref="GI:12862418"
                     /translation="MSQVLIESVETIIVDLPTIRPHKLAMHTMQHQTLVVIRLRCSDG
                     IEGLGESTTIGGLAYGNESPESIKQNIDSHLAPLLIGQDAANINAAMLRLDKAAKGNT
```

```
                FAKSGLESALLDAQGKRLGLPVSELLGGRVRDGLEVAWTLASGDTAKDIAEAQHMLEI
                RRHRIFKLKIGAGEVDRDLKHVIAIKQALGDTASVRVDVNQAWDESVALRACRILGDN
                GIDLIEQPISRINRSGQVRLNQRSPAPIMADESIESVEDAFSLAADGAASIFALKIAK
                NGGPRAVLRTAQIAEAAGIALYGGTMLEGSIGTLASAHAFVTLKQLTWHTELFGPLLL
                TEEIVTEAPVYRDFQLHVPRTPGLGLALDEERLAFFRRK"
     gene       2750. .3040
                /gene="catC"
     CDS        2750. .3040
                /gene="catC"
                /function="muconolactone isomerase"
                /codon_start=1
                /transl_table=11
                /product="muconolactone isomerase CatC"
                /protein_id="BAB32457.1"
                /db_xref="GI:12862419"
                /translation="MLFHVKMTVRLPVDMDPAKAAKLKADEKELAQRLQREGKWRHLW
                RIAGHYANYSVFDLESVEALHDTLMQLPLFPYMEIEVDGLCRHPSSIHEDDR"
     gene       3080. .4009
                /gene="catA"
     CDS        3080. .4009
                /gene="catA"
                /function="catechol 1,2-dioxygenase"
                /codon_start=1
                /transl_table=11
                /product="catechol 1,2-dioxygenase CatA"
                /protein_id="BAB32458.1"
                /db_xref="GI:12862420"
                /translation="MTIKLSSTESVQKFFQEASGFNNDQGSQRMKKLVNRILLDTVKI
                IEDLEVTTEEFWKAVDYLNRMGARGEAGLLVAGLGLEHYLDLLLDAQDEAVGLSGGTP
                RTIEGPLYVAGAPLSEGETRMDDGTDAGTVMFLEGQVVDADGKPIAGATVDLWHANTK
                GTYSYFDTTQSDYNLRRRILTDANGRYRARSIVPSGYGCPPDGTTQEVLNLLGRHGNR
                PAHIHFFISAPGYRHLTTQINLSGDEYLWDDFAYATRDGLVGEVRFTEDAAAAKARGV
                DGSRFAEIEFGFQLQKAPTAEQEAIRDRVRAEA"
     gene       4311. .5180
                /gene="ORFcat5"
     CDS        4311. .5180
                /gene="ORFcat5"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB32459.1"
                /db_xref="GI:12862421"
                /translation="MTRPNEPAELAAYYDVQYNARESVEDFDQYPRQYRALSDDAHAR
                LRCIKDVAYGPGAGERLDIFPAERPDAPVLLFIHGGYWRALSKADSAFMAPALTAAGA
                CVVVLDYDLAPAVTLDHIVDQTRRALAWLHRHIAEFGGDPQRLYASGSSAGGHLTGML
                LAGGWHADYGVPDNVLRGALPISGLFDLRPLLETHINGWMGMDDAAARRNSPSFQLPT
                RGAELVISYGALETAEFARQSHEFLEAWTARGLPGRFVAAPGRNHFDVVLELGQPGTP
                LYQAMIELMGLNL"
     gene       complement(5299. .5748)
                /gene="ORFcat6"
     CDS        complement(5299. .5748)
                /gene="ORFcat6"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB32460.1"
                /db_xref="GI:12862422"
                /translation="MGHITEVFNDLSEQSQRIGSIVGSIQEIAKQTNLLALNAAIEAA
                RAGEQGRGFAVVADEVRNLARRANEASEQIRQIVGGLQQASSDARSGLGQVDGSTRTG
                LEKSDVALEALAQMKFVATARYEIVQRIMQRLDEKQRLAGSLRQLLG"
     gene       complement(6331. .7584)
                /gene="ORFcat7"
     CDS        complement(6331. .7584)
                /gene="ORFcat7"
                /function="channel-forming protein"
                /codon_start=1
                /transl_table=11
                /product="probable channel-forming protein"
                /protein_id="BAB32461.1"
                /db_xref="GI:12862423"
                /translation="MIYKSRALPLLLSALPLCVQAAESGFIEDSTATLQLRNYYFSRD
                FSDIVGPNQQSKAEEWAQGFILNFKSGYTQGPLGFGIDAIGTPRHQARQQPRPQQHRP
                AAGAGGRRARRRLQPPGVAGKMRYSKTELKVGELQPNVPVLVFSDNRLLPPSYQGVSV
                TSGEIAALTLQGGHLNTTSLRNESGRSDIQAIVGYQPQRQAPATPSTYAGGDYAFNAN
                RTTLSAWYGQLEDIYDQRHLGLKHAEPVGDWVLGANLGFYDSNEDGKALIGSIDNQAF
                YLLLSAKRGGNTFFVGYQAMFGDQGFPRVFANISPLGNEVPTYEFASADQRSWQVRHD
                YDFAALGIPGLTTTVRYITGDNVDTGKGFEGKDWERDLDIGYVIQSGSLKGLGIRVRN
                VTARSNYRTDIDENRLILNYTLALF"
     gene       complement(7712. .8401)
                /gene="ORFcat8"
     CDS        complement(7712. .8401)
                /gene="ORFcat8"
                /function="probable ABC transporter subunit"
                /codon_start=1
                /transl_table=11
                /product="probable ABC transporter subunit"
                /protein_id="BAB32462.1"
                /db_xref="GI:12862424"
                /translation="MLEVSGLQAGYGLSQVLFDMHLKIEQGEVVSLIGRNGMGKTTTV
                KSIMGLLKPSAGSIRIHGKEMRGALPFRIAQAGLGLVPEGRRAFGSLSVKENLLATAT
                RGGKWDLARIYKLFPRLEERQEQLSRTLSGGEQQMLVVARALMTNPQLLILDEATEGL
                APVIRETIWNCLATLKAEGETILVIDKNLKEMARVVDRHHIIEKGRQVWQGTPEQLAA
                TPELSHRYLGV"
BASE COUNT     1487 a   2919 c   3001 g   1511 t
ORIGIN


  Query Match             2.9%;  Score 38.8;  DB 1;  Length 8918;
  Best Local Similarity   53.7%;  Pred. No. 65;
  Matches  102;  Conservative    0;  Mismatches   87;  Indels    1;  Gaps    1;

Qy   1095 gacagccgtcaacggcggcgaccgcgccatggtaccg-atcggcacttatgagcgcgtaa 1153
          | | || ||   || |||||  ||||||  |  | | | | |||||      |    |
Db   7232 GGCGGCTGTAGTCGTCGGCGGGCGCGCCGTCCTCCTGCACCGGCAGCAGGCCGGTGTTGC 7291

Qy   1154 tgccgttggacatcctgcctaccttgcttttgcgcgatttaatcgtcggcgataccgaca 1213
          ||| || ||   | ||| | | |||| |   | | | |      | || |||||||||
Db   7292 TGCGGTCGGGGCTGCTGTCGAGCTTGATGCCGAGGGGTGCCGATGGCGTCGATACCGAAG 7351

Qy   1214 gcgcgcaggctttgggttgcttggaattggacgaagaagacctcgctttgtgcagcttcg 1273
           |  || |||  |||||      | | | || |  || ||    ||  || ||   |||
Db   7352 CCCAGCGGGCCCTGGGTGTAGCCGGACTTGAAGTTGAGGATGAAGCCCTGGGCCCATTCC 7411

Qy   1274 tctgcccggg 1283
          || |||  ||
Db   7412 TCGGCCTTGG 7421


RESULT  29
AF263912
LOCUS       AF263912              123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
```

```
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
     gene            4714. .5748
                     /gene="nysDIII"
     CDS             4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
     gene            5930. .34363
                     /gene="nysI"
     CDS             5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                     AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                     HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                     ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                     DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                     ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                     DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                     VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                     GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                     AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                     LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                     LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                     QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                     LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                     DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                     SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                     PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                     AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                     SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                     LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                     SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                     ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                     VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                     PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                     VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                     TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                     PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                     CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEEELAAATALAR
                     LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                     VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                     TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
```

```
ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
TAGRPAGTWTPRGTTLVTGGSGTLAPHLARWLAQRGAEHLVLISRRGTAAPGAAELVA
```

Query Match 2.9%; Score 38.8; DB 1; Length 123580;
Best Local Similarity 48.6%; Pred. No. 79;
Matches 106; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy   1104 caacggcggcgaccgcgccatggtaccgatcggcacttatgagcgcgtaatgccgttgga 1163
          || || ||||| || || |   |   | ||||||       || ||    || |
Db  49632 CACCGCCGGCGTCCTCGACGACGCCACCGTCGGCACCCTGACGCCCGAGCGGCTGGACAC 49691

Qy   1164 catcctgcctaccttgcttttgcgcgatttaatcgtcggcgataccgacagcgcgcaggc 1223
          | ||||||   ||  |        ||  |    | |   ||| ||  | ||| |  |
Db  49692 CGTCCTGCGCGCCAAGGCGGACGCCGCCTGGCACCTGCACGACGCCACCCGCGACCGCGA 49751

Qy   1224 tttgggttgcttggaattggacgaagaagacctcgctttgtgcagcttcgtctgcccggg 1283
            |||   | || |   || ||      | |  || | |   | ||  |  | ||| |||
Db  49752 CCTGGCAGGGTTCGTGCTGTACTCCTCGGTCGCCGGTGTCACCGGCGGCCCCGGCCAGGG 49811

Qy   1284 caaatacgaatacggcccgctgttgcgcaaagtgctgg 1321
          ||| ||||    ||||     ||| | | | | ||| |
Db  49812 CAACTACGCCGCCGGCAACACGTTCCTCGACGCGCTCG 49849
```

```
RESULT 30
AX211739
LOCUS       AX211739               125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN
```

Query Match 2.9%; Score 38.8; DB 6; Length 125401;
Best Local Similarity 48.6%; Pred. No. 79;
Matches 106; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy   1104 caacggcggcgaccgcgccatggtaccgatcggcacttatgagcgcgtaatgccgttgga 1163
          || || ||||| || || |   |   | ||||||       || ||    || |
Db  50040 CACCGCCGGCGTCCTCGACGACGCCACCGTCGGCACCCTGACGCCCGAGCGGCTGGACAC 50099

Qy   1164 catcctgcctaccttgcttttgcgcgatttaatcgtcggcgataccgacagcgcgcaggc 1223
          | ||||||   ||  |        ||  |    | |   ||| ||  | ||| |  |
Db  50100 CGTCCTGCGCGCCAAGGCGGACGCCGCCTGGCACCTGCACGACGCCACCCGCGACCGCGA 50159

Qy   1224 tttgggttgcttggaattggacgaagaagacctcgctttgtgcagcttcgtctgcccggg 1283
            |||   | || |   || ||      | |  || | |   | ||  |  | ||| |||
Db  50160 CCTGGCAGGGTTCGTGCTGTACTCCTCGGTCGCCGGTGTCACCGGCGGCCCCGGCCAGGG 50219

Qy   1284 caaatacgaatacggcccgctgttgcgcaaagtgctgg 1321
          ||| ||||    ||||     ||| | | | | ||| |
Db  50220 CAACTACGCCGCCGGCAACACGTTCCTCGACGCGCTCG 50257
```

```
RESULT 31
AE005062
LOCUS       AE005062               10500 bp    DNA     linear   BCT 12-FEB-2001
DEFINITION  Halobacterium sp. NRC-1 section 93 of 170 of the complete genome.
ACCESSION   AE005062 AE004437
VERSION     AE005062.1  GI:10580956
KEYWORDS    .
SOURCE      Halobacterium sp. NRC-1.
  ORGANISM  Halobacterium sp. NRC-1
            Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
            Halobacteriaceae; Halobacterium.
REFERENCE   1  (bases 1 to 10500)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     From the cover: genome sequence of halobacterium species NRC-1
  JOURNAL   Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)
   PUBMED   11016950
REFERENCE   2  (bases 1 to 10500)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Institute for Systems Biology, 4225
            Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES             Location/Qualifiers
     source          1. .10500
                     /organism="Halobacterium sp. NRC-1"
```

```
                /strain="NRC-1"
                /db_xref="taxon:64091"
     gene       complement(86. .232)
                /gene="VNG1459H"
     CDS        complement(86. .232)
                /gene="VNG1459H"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /product="Vng1459h"
                /protein_id="AAG19765.1"
                /db_xref="GI:10580957"
                /translation="MLDLSVRAADDKGMGVCPECHGPVEKIDPWFGATRIECRRCGEV
                YHEY"
     gene       complement(522. .734)
                /gene="VNG1461H"
     CDS        complement(522. .734)
                /gene="VNG1461H"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /product="Vng1461h"
                /protein_id="AAG19766.1"
                /db_xref="GI:10580958"
                /translation="MVWHIVEDAAGRSGGMEKQDLVVAVHVMVAVAIAAFGLVRISRG
                QRVPGALNVGFAIVVVGVGVYMRQLV"
     gene       complement(734. .2809)
                /gene="cdc48a"
                /note="VNG1462G"
     CDS        complement(734. .2809)
                /gene="cdc48a"
                /note="Cdc48a"
                /codon_start=1
                /transl_table=11
                /product="cell division cycle protein"
                /protein_id="AAG19767.1"
                /db_xref="GI:10580959"
                /translation="MTRVRVTLRDAADRTADAPAVGPGQLSFTAETRQQIGVAEGDAV
                TVTGSEDTVATVGPDTTAIAASSGLAGADVLRNVGPTDTVTVAPADPSPAAELTVAPA
                TSSTLEGDADALARALAGRHLVADDRIAVPLFSGTVVVTLDVEDTTPGGPVTVTADTA
                ITLTGHAADSGHADAEHATPADTRVGGLDDERGALRRLVVAPLVADSYAAIGVRPPAG
                VLVHGPAGTGKTTLVRAVAAAADLAVESVAPEDAGDRDALAAVLDAARDAEPGCVVFV
                ESLAAAAPDPTADGASGRGSPSALGWLLDRVRGHDTVVVVGETTDPDAVDPALRRGGR
                FDAEVRVGVPDPAARRAILDVHTDGVRLADAVSLDAVADRTHGYTGADLTAVLVDAAT
                RAAGSAAGPPVIRQRDLEAALDAVGPSTLRDASVQTPTTTYQDIGGLDRAKREVVRTV
                EWPQRYPALFERLDAAAPTGVLLHGPPGTGKTMLAKAVAASTDANFLSVDGPELMNRY
                VGESERGVRDLFERARRLAPAVVFLDEVDSLAPARHDTDTGASERVVSQLLTELDGLS
                PRGSVAVLAATNRRESVDPALLRPGRIETQVAVPIPDQDARAAIFEVQLDGVATGRID
                TTALAAATTGYTGSDIAGVVREGALLAMEDHLRETEFDATDASGLVVTQRHLRRAIET
                TAPSVDGAG"
     gene       3000. .3428
                /gene="blp"
                /note="VNG1463G"
     CDS        3000. .3428
                /gene="blp"
                /note="Blp"
                /codon_start=1
                /transl_table=11
                /product="bacterio-opsin linked product"
                /protein_id="AAG19768.1"
                /db_xref="GI:10580960"
                /translation="MFTVEQQATIDAPPGDVFSFLDDPRNHVKITPGLITVTDVEALA
                SGGKRAKYRYKLAGVELVGTVEDIERVPERRLVQELTGAIDGTITYRLSGDDTTSLTY
                EAEYDLPDTVIESVLAPIAKAYNKREAAATVDNLTTFLEH"
     gene       complement(3462. .5486)
                /gene="bat"
                /note="VNG1464G"
     CDS        complement(3462. .5486)
                /gene="bat"
                /note="Bat"
                /codon_start=1
                /transl_table=11
                /product="bacterio-opsin activator"
                /protein_id="AAG19769.1"
                /db_xref="GI:10580961"
                /translation="MTSVQNTESETAAGATTIGVLFAGSDPETGPAACDLDEDGRFDV
                TQIRDFVAARDRVDDPDIDCVVAVHEPDGFDGVAFLEAVRQTHAEFPVVVVPTAVDED
                VARRAVDADATGLVPAVSEDATAAIADRIEQSAPAHSEDTETRMPISDLTVESERRLK
                EQALDEAPIGITISDATDPEEPIIYINDSFEDITGYSPDEVVGANHRFLQGPKTNEDR
                VAEFWTAITEDHDTQVVLRNYRRDGSLFWNQVDISPIYDEDGTVSHYVGFQMDVSERM
                AAQQELQGERQSLDRLLDRVNGLMNDVTSALVRAADREEIETRITDRIGTGGEYAGAW
                FGRYDATEDTITVAEAAGDCEGCDGDVFDLASAGEAVALLQDVVEQREALVSTDADGV
                SGTADGDACVLVPVTYRSTTYGVLAVSTAEHRIDDREQVLLRSLGRTTGASINDALTR
                RTIATDTVLNIGVELSDTALFLVELAGATDTTFEQEATIADSQTQGVLMLVTTPHDDP
                QAVVDTALGYDAVQDAEVIVSTDDESVVQFDLSSSPLVDVLSECGSRVIRMHADRTTL
                ELDVRVGTEGAARRVLSTLRDKYADVELVAYHEDDPEQTPHGFREELRNDLTDRQLTA
                LQKAYVSGYFEWPRRAEGKQLAESMDIVPSTYHQHLQAAKQKLVGAFFEE"
     gene       complement(5483. .6589)
                /gene="brp"
                /note="VNG1465G"
     CDS        complement(5483. .6589)
                /gene="brp"
                /note="Brp"
                /codon_start=1
                /transl_table=11
                /product="bacteriorhodopsin related protein"
                /protein_id="AAG19770.1"
                /db_xref="GI:10580962"
                /translation="MLGSDVCIHMSNRSQFVPSWLVPEAAGDLPLTVSRLSLLALAAA
                FAVGYGAGFAVPLEVQAGVYLLGMVAMNLPHGGYEHFENLRRRAASFQGKYIVAYLVG
                IAAFGALFFVAPVAGLGLAVTVAVAKGGFGGVQSMDALYGTDHLRTRPQRWLAAVVRG
                GAVMVVPMLFWTDVFYAFSSVMISIFDPSAVSALGGDIATRRLVLGGGYGALVVAHLG
                LGYRRAAGTGSFLADAAETLLLIAYFALVPVVIAVGLYFPLWYSARQVARSSAVDDTA
                MTQADATGMLDALDADDPARATLASWAVLIVGSVATFGLAAVLWLLSPQPLGGGGILV
                GLVAFWSIFVSIIALPHVVVGGWLDRTRGIWYVP"
     gene       6711. .6845
                /gene="VNG1466H"
     CDS        6711. .6845
                /gene="VNG1466H"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /product="Vng1466h"
                /protein_id="AAG19771.1"
                /db_xref="GI:10580963"
                /translation="MGLPMGATVKSVTAASRQEPTSDTQKVRTVECRNDHEFFVRFEW
                "
     gene       7089. .7877
                /gene="bop"
                /note="VNG1467G"
     CDS        7089. .7877
                /gene="bop"
                /note="Bop"
                /codon_start=1
                /transl_table=11
                /product="bacteriorhodopsin"
                /protein_id="AAG19772.1"
                /db_xref="GI:10580964"
                /translation="MLELLPTAVEGVSQAQITGRPEWIWLALGTALMGLGTLYFLVKG
                MGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVPFGGEQNPIYWARYADWLFT
                TPLLLLDLALLVDADQGTILALVGADGIMIGTGLVGALTKVYSYRFVWWAISTAAMLY
                ILYVLFFGFTSKAESMRPEVASTFKVLRNVTVVLWSAYPVVWLIGSEGAGIVPLNIET
                LLFMVLDVSAKVGFGLILLRSRAIFGEAEAPEPSAGDGAAATSD"
     gene       7929. .8198
                /gene="VNG1468H"
     CDS        7929. .8198
                /gene="VNG1468H"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /product="Vng1468h"
                /protein_id="AAG19773.1"
                /db_xref="GI:10580965"
                /translation="MSPPLGLVLARSRDDGDADGDFPASLNRLAVVALAGAALVGAAG
                LFAVPFLRSFGMTFWEAFTVVGVSEFVSAIVAALAGYHLYTTPDD"
     gene       8329. .9603
```

```
                     /gene="pri"
                     /note="VNG1470G"
     CDS             8329. .9603
                     /gene="pri"
                     /note="Pri"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA primase"
                     /protein_id="AAG19774.1"
                     /db_xref="GI:10580966"
                     /translation="MHRAPGRFVRPRRAVDPGETTLIYPRSAYCRPDMEERTRAYLRG
                     RFGDVYRRAEIDLPPRADDREWGYIPWTAGPDTTMVRHKSTLDLGSMTSFLERKCPRH
                     VYFSAGTYDAPGAATMDEKHWQGSDLVFDLDADHLPRVTLGEDSYAEMLSKCKDALLR
                     LLDFLERDFGFEELTVTFSGGRGYHVHVRDAGVYELGSEERREIVDYVRGNDLALETL
                     VEAEPVGGRGLENPTEKRMLPADGGWGRRVTTRLEAFADDLIERGEDDAVATLTEFDG
                     VGERSARAIYNVVADNTTAVKRGNVDVHPAFLTVARRYIDETVAAEQAPIDEPVTTDT

  Query Match              2.9%;  Score 38.4;  DB 1;  Length 10500;
  Best Local Similarity   52.5%;  Pred. No. 83;
  Matches   84;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     397 acccgtccgttcagcaaaatccctgccgtagatgccgagccgttcgccatcttcgtcaat 456
           |||    |  || |  | ||||  | ||| || |     || | || | ||   | |
Db   10293 ACCGACACCGTCGGTGAGATCCTCGGCGTCGACGAGCGACCCTACGACCTCGAAGCCGGC 10352

Qy     457 gcgatggacaccaatccgctggctgccgaccctacggtcatcatcaaagaagccgccgaa 516
           |   ||| |||    |||  |||   || |    || |||||   ||||| |||||||| 
Db   10353 GACGTGGTCACGCTCCCGGAGGCGAACGCCGACCCGCTCATCGAGAAAGACGCCGCCGAG 10412

Qy     517 gacttcaaacgcggcctgttggtattgagccgcctgaccg 556
           |  |  |  | |||| |  ||      |||  ||| | |
Db   10413 AAGCTGGACTGAGGCCGGGCGGGCGACGGCCTGCTGTCGG 10452


RESULT  32
AF220510
LOCUS       AF220510               3000 bp    DNA     linear   BCT 07-MAR-2001
DEFINITION  Azoarcus evansii BoxB (boxB) and BoxA (boxA) genes, complete cds.
ACCESSION   AF220510
VERSION     AF220510.1  GI:12655803
KEYWORDS    .
SOURCE      Azoarcus evansii.
  ORGANISM  Azoarcus evansii
            Bacteria; Proteobacteria; beta subdivision; Rhodocyclus group;
            Azoarcus.
REFERENCE   1  (bases 1 to 3000)
  AUTHORS   Mohamed,M.E., Zaar,A., Ebenau-Jehle,C. and Fuchs,G.
  TITLE     Reinvestigation of a New Type of Aerobic Benzoate Metabolism in the
            Proteobacterium Azoarcus evansii
  JOURNAL   J. Bacteriol. 183 (6), 1899-1908 (2001)
   PUBMED   11222587
REFERENCE   2  (bases 1 to 3000)
  AUTHORS   Mohamed,M.E. and Fuchs,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2000) Mikrobiologie, Inst. f. Biologie 2,
            Schaenzlestr. 1, Freiburg 79104, Germany
FEATURES             Location/Qualifiers
     source          1. .3000
                     /organism="Azoarcus evansii"
                     /strain="KB740"
                     /db_xref="taxon:59406"
     gene            108. .1529
                     /gene="boxB"
     CDS             108. .1529
                     /gene="boxB"
                     /function="involved in aerobic benzoate metabolism"
                     /note="benzoyl-CoA oxygenase component"
                     /codon_start=1
                     /transl_table=11
                     /product="BoxB"
                     /protein_id="AAK00599.1"
                     /db_xref="GI:12655804"
                     /translation="MINYSERIPNNVNLNENKTLQRALEQWQPSFLNWWDDMGPENSS
                     NYDVYLRTAVSVDPKGWADFGYVKMHDYRWGIFLAPQEGEKKITFGEHKGQDVWQEVP
                     GEYRSTLRRIIVTQGDTEPASVEQQRHLGLTAPSLYDLRNLFQVNVEEGRHLWAMVYL
                     LHAHFGRDGREEGEALLERRSGDEDNPRILTAFNEKTPDWLSFFMFTFITDRDGKFQL
                     ASLAESAFDPLARTCKFMLTEEAHHLFVGESGIARVIQRTCEVMKELGTDDPAKLRAA
                     GVIDLPTLQKYLNFHYSVTSDLYGAEISSNAATYYTNGLKGRFEEEKIGDDHKLQNSE
                     YEVMDVAGDKILTRHVPALSALNERLRDDWITDVQAGVDRWNRIPAKFGFDFRFTLPH
                     KGFHRKIGMFADVHVSPDGRLISEAEWTHQHKNWLPTESDRLYVHSLMGRCLEPGKFA
                     NWIAAPARGINNQPVNFEYVRFN"
     gene            1692. .2936
                     /gene="boxA"
     CDS             1692. .2936
                     /gene="boxA"
                     /function="involved in aerobic benzoate metabolism"
                     /note="benzoyl-CoA oxygenase component"
                     /codon_start=1
                     /transl_table=11
                     /product="BoxA"
                     /protein_id="AAK00600.1"
                     /db_xref="GI:12655805"
                     /translation="MNAPAEHANLARQHLIDPEICIRCNTCEEICPVDAITHDSRNYV
                     VKFETCNGCLACISPCPTGAIDSWRNVDKATPHSLADQYSWDYLPDTTELDQFEATVM
                     GAAELPAEVQQITEVATAGQGGPAMAPWSASHPYVNLYTPANPITATVTGNYRLTAED
                     ASSDIHHIVLDFGTTPFPVLEGQSIGIIPPGVDEKGKPHLLRMYSVASPRDGERPHYN
                     NLSLTVKRVVEDHEGNPTRGVASNYVCDLKKGDKVQVTGPYGSTYLMPNHPGSSIMMI
                     CTGTGSAPMRAMTERRRRRMDRKEGGELVLFFGARAPEELPYFGPLQKLPKEFIDINF
                     AFSRVPGEPKRYVQDAIRERADKVFQMLQDDNCYIYICGLKGMEAGVLEAFRDICRAK
                     GADWDALRPQLLSKARFHVETY"
BASE COUNT      602 a   1089 c    883 g    426 t
ORIGIN


  Query Match              2.8%;  Score 38.2;  DB 1;  Length 3000;
  Best Local Similarity   49.7%;  Pred. No. 85;
  Matches   97;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy     405 gttcagcaaaatccctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatgga 464
           ||||| |   ||| | | |   || | | ||        |  ||   ||   |    ||
Db     722 GTTCACCTTCATCACCGACCGTGACGGCAAGTTCCAGCTCGCCTCGCTCGCCGAATCGGC 781

Qy     465 caccaatccgctggctgccgaccctacggtcatcatcaaagaagccgccgaagacttcaa 524
           |  | | ||||||||    |  |   | | ||||  |||  || |  ||  |  || ||
Db     782 CTTCGACCCGCTGGCGCGCACCTGCAAGTTCATGCTCACCGAGGAGGCGCACCACCTCTT 841

Qy     525 acgcggcctgttggtattgagccgcctgaccgaacgtaaaatccatgtgtgtaaagcagc 584
           ||||   | ||   |   |||| ||| | | || |    | | |||   || | |
Db     842 CGTCGGCGAATCGGGCATCGCCCGCGTGATCCAGCGCACCTGCGAAGTGATGAAGGAACT 901

Qy     585 aggcgcagacgtgcc 599
           ||| | ||||  ||
Db     902 CGGCACCGACGACCC 916


RESULT  33
AE004915/c
LOCUS       AE004915               9980 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 476 of 529 of the complete
            genome.
ACCESSION   AE004915 AE004091
VERSION     AE004915.1  GI:9951308
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 9980)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
```

```
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 9980)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .9980
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            240. .1856
                     /gene="PA5021"
     CDS             240. .1856
                     /gene="PA5021"
                     /codon_start=1
                     /transl_table=11
                     /product="probable sodium/hydrogen antiporter"
                     /protein_id="AAG08406.1"
                     /db_xref="GI:9951309"
                     /translation="MIAGNDGVGGIVFDNYPMAYLVGNLALAVILLDGGLRTRVSSFR
                     VALWPALSLATLGVLVTTGLTGIAAAWLFDLHWMEGLLIGAIVGSTDAAAVFSLLGGK
                     GLNERVTATLEIESGSNDPMAVFLTVTLIEMLASGQTGLSWGFALHLVQQFGLGALLG
                     LGGGWLLLQLINRMHLAGGLYPLLVISGGLLVFALANAVGGSGILAIYLCGLLLGNRP
                     IRSRHGILHMLDGMAWLAQIGMFLVLGLLVTPHDLWPIALPALALALWMILVARPLSV
                     LIGLIPFRAFHDREKAFIAWVGLRGAVPIILAVFPLMAGLPNAQLFFNVAFFIVLVSL
                     LVQGTSLPWAARLLRVVVPPDPAPISRAGLEIHPTSEWELFVYHLNKEKWCIGAALRE
                     LKMPGGTRIAALFRGTELLHPSGSTILEADDILCVIGHEHDLPALGKLFSQAPDRGLG
                     ARFFGDFVLEGDAQLSAVASLYGLKLDGIDGEQALGRFIAHEIGGEAVIGDQVEWNGL
                     TWTVAALEGNRIRKVGVKFPEGRPGPGLFL"
     gene            1948. .5304
                     /gene="PA5022"
     CDS             1948. .5304
                     /gene="PA5022"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG08407.1"
                     /db_xref="GI:9951310"
                     /translation="MPALRSLIAILFLGLCLASPPLLAQSEPPSAEAVQQSLDKLAER
                     KLAEADQKVAKASLEQTLKFLAARDEALQSLEDLKKRLSDAPRQIEENQRELERLKKT
                     KERPVSERYSGESAARLEMLLNDRTTQQAEWQKALGEANSLSITAETRPERAQAGISS
                     MQARILEIGSLLKAGKESGKTINADRRGELLAEQAALTVQSQLLRQELAGNNLLQDLG
                     KSQHDLLTEKISRLEKETLDLQALISEKRREQSEKTVAELSKEGAQGAGTDSLLSQEN
                     AKNLRLSDYLLRATDRLNVLTRRNLETKQQLDNLTQSNQALEEQINVLRGSLLLSRIL
                     YKQKQALPKIKADQSLADEIADLRLGQFELNQERDKLATPQQYLDDLLAQQPSEQVTP
                     ELRKDLDTLLATRSELLERLNHELNALLNEAITLQLNQKQLLSTSESLRTTLDEQMFW
                     IPSNQPLDLSWFKMTPTLLKNQLTEIPWGSGVRELGEGLVDRPLLFLPLFLLIAALLW
                     KRRYLYDKLAELNDDIGHFKRDSQLHTPLAILITILLALPGTLTFAVAGLALLLDARG
                     QNATLGSALLEMAQVWLLFYTAHRILTPGGIAERHFHWSREQVGFLDRHLVRLGVVVL
                     ALVGVVTVAEHQPASLADDVIGIAVVLLGYAAMAWLLSRLLFSSPGDERPSLIKMIVG
                     ILFTALPLVLFVAVCFGYYYTSLKLTDRLIDTLYLLLLWFVIEAAFVRGLGVAARRLA
                     YARALAKRQNAAKEGVDGETNVEEPTLGIEQINQQSLRLIRLALLIGFIVCLYWVWAD
                     LISVISYLDNVTLYQYTAGTGEAATSVPISLRDFLAALAIAVVTVALARNLPGLLEVL
                     VLQRLTLAQGSAYATTTLLSYAISGIGIVSALSTLGVSWDKLQWLVAALSVGLGFGLQ
                     AIFSNFVSGLIILFERPVRIGDVVTIGNLSGTVSRIRIRATTITDFDRKEIIVPNQTF
                     ITGQLVNWSLSDTVTRVTIKIGLAYETDLPLARKLMMQAVKENPRVLRDPEPLLYFLT
                     ISASTFDYELRFHVRELGDRNPSTDEILTKIATSFREHKIDMAFNQVDVFVKNLQGQQ
                     VQLFVGEPQHSGVGNPPALKGQDGPAEPTR"
     gene            5371. .6831
                     /gene="PA5023"
     CDS             5371. .6831
                     /gene="PA5023"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG08408.1"
                     /db_xref="GI:9951311"
                     /translation="MKSLDDLDFDNRFARLGGAFSTEVLPDPIAEPRLVVASPAALAL
                     LDLPAETSDEALFAELFGGHKLWSEAEPRAMVYSGHQFGSYNPRLGDGRGLLLGEVIN
                     QAGEHWDLHLKGAGQTPYSRMGDGRAVLRSSIREFLASEALPALGIPSSRALCVIGSS
                     TPVWREKKESAATLLRLAPSHVRFGHFEYFYYTRQHDQLKQLAAFVLEHHFADCNAAE
                     RPYAAMFRQVVERNAELIARWQAYGFCHGVMNTDNMSILGITFDYGPYAFLDDFDANH
                     ICNHSDDAGRYSFSNQVPIAHWNLAALAQALTPLVEVDELRASLDLFLPLYQAHYLDL
                     MRRRLGLGVAAENDHALVQELLQRMQGSAVDYSLFFRRLGEETPERALASLRDDFVDR
                     EAFDRWAEAYRRRVEEEGGDQESRRRRMHAVNPLYVLRNYLAQQAIEAAEQGDYTEVR
                     LLHQVLSRPFEEQPGMERFTRRPPDWGRHLEISCSS"
     gene            7237. .7953
                     /gene="PA5024"
     CDS             7237. .7953
                     /gene="PA5024"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG08409.1"
                     /db_xref="GI:9951312"
                     /translation="MAVGFFAQVIDGALGMAYGITATTFLLSAGATPAAASASVHIAE
                     VFTTGFSGISHVKLGNVNKSLFLRLLLPGMIGAVLGAVLITRFDGHQLKPFISAYLLL
                     MGLYILSKAYRHVIKRRAPRHVAKLALFGGFVDAAGGGWGPVVTSSLIGSGSDPRTTI
                     GSVNFAEFFLTIASATSFILLAGQPDTWKMVAGLVFGGLFAAPFAALLCKKLSPRTLL
                     VIVGTLITLISAYNIYIGLR"
     gene            8074. .9351
                     /gene="metY"
                     /note="PA5025"
     CDS             8074. .9351
                     /gene="metY"
                     /codon_start=1
                     /transl_table=11
                     /product="homocysteine synthase"
                     /protein_id="AAG08410.1"
                     /db_xref="GI:9951313"
                     /translation="MKLETLAVHAGYSPDPTTRAVAVPIYQTTSYAFDDTQHGADLFD
                     LKVPGNIYTRIMNPTNDVLEQRVAALEGGVGALAVASGMAAITYAIQTVAEAGDNIVS
                     VAKLYGGTYNLLAHTLPRIGIQARFAAHDDVAALEALIDERTKAVFCETIGNPAGNII
                     DLQALADAAHRHGVPLIVDNTVATPVLCRPFEHGADIVVHSLTKYMGGHGTSIGGIVV
                     DSGKFDWAANKSRFPLLNTPDPSYHGVTYTEAFGPAAFIGRCRVVPLRNMGAALSPFN
                     AFLILQGLETLALRMERHCDNALAVARYLQQHPQVAWVKYAGLADNPEHALARRYLGG
                     RPAAILSFGIQGGSAAGARFIDALKLVVRLVNIGDAKSLACHPASTTHRQLNAEELAR
                     AGVSDDMVRLSIGIEHIDDILADLDQALAAAAR"
     gene            9442. .9894
                     /gene="PA5026"
     CDS             9442. .9894
                     /gene="PA5026"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG08411.1"
                     /db_xref="GI:9951314"
                     /translation="MPLPTELARHLTEEKIAFVQRSGLRAEVLEPGYVRLRMPGAGNE
                     NHIGSMYAGALFTLAELPGGALFLTSFDSARFYPIVKEMTLRFRRPAKGDIRVEARLD
                     AERIRQLETEAGERGKAEYSLELQLTDEQGEVVAESAALYQLRSHARP"
BASE COUNT     1697 a   3648 c   3006 g   1629 t
ORIGIN


  Query Match             2.8%;  Score 38.2;  DB 1;  Length 9980;
  Best Local Similarity   55.7%;  Pred. No. 93;
  Matches   73;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     417 ccctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgct 476
           ||  ||||  | ||  | || | |||| | |   | | ||  |   |      | |  |
Db    7723 CCGAGCCGATGGTGGTGCGCGGGTCGCTACCCGAGCCGATCAGGCTGCTGGTGACCACCG 7664

Qy     477 ggctgccgaccctacggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgtt 536
           |||  | | | |  | | || |||| | ||||||||||||||| | | | |  ||||   |
Db    7663 GGCCCCAGCCGCCGCCGGCAGCATCGACGAAGCCGCCGAAGAGTGCCAGCTTGGCCACGT 7604

Qy     537 ggtattgagcc 547
           ||    | |||
Db    7603 CCATAACTCGG 7593
```

```
Db   7603 GGCGCGGTGCC 7593


RESULT  34
SC10A5
LOCUS       SC10A5                  38734 bp    DNA     linear   BCT 26-APR-2000
DEFINITION  Streptomyces coelicolor cosmid 10A5.
ACCESSION   AL021529
VERSION     AL021529.1  GI:2815309
KEYWORDS    ABC-transporter; cold-shock domain protein; DEAD-box RNA helicase;
            NTP pyrophosphohydrolase; oxidase; oxidoreductase; periplasmic
            oligopeptide binding protein; phosphoenolpyruvate-utilizing enzyme;
            secreted protease; transcriptional regulator; transposase
            pseudogene; xlnA; xylanase A.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 38734)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 38734)
  AUTHORS   Murphy,L. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 38734)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
            program of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid 10A5 lies between St3F7 and St7H1 on the AseI-B genomic
            restriction fragment.
FEATURES             Location/Qualifiers
     source          1. .38734
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 10A5"
     gene            1. .289
                     /gene="SC10A5.01"
     misc_feature    1. .108
                     /gene="SC10A5.01"
                     /note="overlap with St3F7 from: 23623 to: 23730"
     CDS             <1. .289
                     /gene="SC10A5.01"
                     /note="SC10A5.01, probable phosphoenolpyruvate-utilizing
                     en zyme, partial CDS, len: >95 aa; overlaps and extends
                     ORF fr om neighbouring cosmid SC3F7.16; similar to many
                     eg. PPSA_S TAMA P46893 probable phosphoenolpyruvate
                     synthase (834 aa), fasta scores; opt: 215 z-score: 351.4
                     E(): 2.3e-12, 47.8% identity in 69 aa overlap"
                     /codon_start=2
                     /transl_table=11
                     /label=SC10A5.01
                     /protein_id="CAA16433.1"
                     /db_xref="GI:2815310"
                     /db_xref="SPTREMBL:O54094"
                     /translation="ILVTGYTDASWTPLFAIAGGVVTDIGSMLSHSSIVAREFHVPSV
                     VNTKDATQRINTGDLIVVDGDAGTVEVVESADTDPQGPAGAAGTPAGATTD"
     RBS             355. .358
                     /note="possible RBS upstream of SC10A5.02"
     gene            367. .1554
                     /gene="SC10A5.02"
     CDS             367. .1554
                     /gene="SC10A5.02"
                     /note="SC10A5.02, probable oxidase, len: 395; similar in
                     N- terminus to many diverse oxidases eg. TR:P95483
                     (EMBL:U7449 3) AMINOPYRROLNITRIN OXIDASE PRND from
                     PSEUDOMONAS FLUORESC ENS (363 aa), fasta scores; opt: 349
                     z-score: 372.1 E(): 1. 6e-13, 33.3% identity in 165 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /label=SC10A5.02
                     /protein_id="CAA16434.1"
                     /db_xref="GI:2815311"
                     /db_xref="SPTREMBL:O54095"
                     /translation="MIPNQWYPIVEAQEVGNDKPLGVRRMGQDLVLWRDIDGNLVCQG
                     ARCPHKGANLGDGRMKGNTIECPYHGFRYGADGACRVIPAMGSEARIPGSLRVPTYPV
                     REQFGLVWMWWGDERPTADLPPVAAPAEVTDNRKLYATKRWTRPVHYTRYIESLLEFY
                     HVTYVHRDHWFNYIDYLLLYGTPSKFGLDGRERYLAATRITNHRVETEAEGQTIRYSF
                     DHCQEDDPTNTTHYVITFTFPCMVHVQTEQFETTSWLVPIDDQNTEHILRWYEYEQVK
                     PVLRFEPLRRLLPWASLYMEKWVQDPQDVRIMEHQEPKISAGGVNKFIPVDEMNAKYI
                     SMRAKLIADASAAPSSPARAAEPEPEAAGRGGSAARATGNGRGAAGGRRGTKPKEDAA
                     ARP"
     gene            1597. .2277
                     /gene="SC10A5.03"
     CDS             1597. .2277
                     /gene="SC10A5.03"
                     /note="SC10A5.03, probable membrane protein, len: 226 aa"
                     /codon_start=1
                     /transl_table=11
                     /label=SC10A5.03
                     /protein_id="CAA16435.1"
                     /db_xref="GI:2815312"
                     /db_xref="SPTREMBL:O54096"
                     /translation="MYGGYDASTGPKALVTAFNTVAVAGAVWFLFGGADTVADWFGTD
                     FDEAVTLRRVLLATLSVLYLLRFIATNFVMLQRKMEWSESATIGIWVLVIHGTMAYFG
                     GTNDAGVSAFTWLGVVLYLLGSYLNTGSEYQRKLWKKRPENKGKLYTEGLFKHSMHIN
                     YFGDAVLFSGFALVTGTPWAFAIPLIMVCMFVFLNIPMLDKYLAERYGEAFDEYASRT
                     AKFVPYVY"
     RBS             2433. .2436
                     /note="possible RBS upstream of SC10A5.04"
     gene            2442. .3329
                     /gene="SC10A5.04"
     CDS             2442. .3329
                     /gene="SC10A5.04"
                     /note="SC10A5.04, unknown, len: 295 aa; similar to
                     hypothet ical proteins from Mycobacterioum tuberculosis
```

and M. lepra e eg. YV29_MYCTU Q11167 hypothetical 39.3 kd protein cy20g9 .29 (358 aa), fasta scores; opt: 1075 z-score: 1514.7 E(): 0, 57.8% identity in 263 aa overlap"
/codon_start=1
/transl_table=11
/label=SC10A5.04
/protein_id="CAA16436.1"
/db_xref="GI:2815313"
/db_xref="SPTREMBL:O54097"
/translation="MTGDIPFGEAEASLTAEVLREVLAGGAEAFARLTSDEGAVDDFG
FDPELTDDYLLPALRLLYEKYFRVDLEGLENVPAEGGALLVANHSGTLPLDALMLQVA
LHDHHSTHRRLRLLAADLAFDLPVVRDLARKAGHVRACPENALRLLGSGELVGVMPEG
YKGLGKPFEERYRLQRFGRGGFAAVALRSRRPMVPCSIVGAEEIYPMIGSAPTLARML
KLPYFPITPTFPLLGALGLIPMPTKWTIRFGAPIHTDGFPEDAAEDPLVVEKLAGEVK
DTIQHTLNEMLEGRGSPFV"

gene 3698. .5773
/gene="SC10A5.05"

CDS 3698. .5773
/gene="SC10A5.05"
/note="SC10A5.05, probable membrane protein, len: 691; high ly similar to M. tuberculosis putative membrane protein TR: O07192 (EMBL:Z96072) MTCY05A6.11C (647 aa), fasta scores; o pt: 1980 z-score: 3557.6 E(): 0, 62.7% identity in 659 aa o verlap and weakly similar to many transport proteins eg. PR OY_SALTY P37460 proline-specific permease proy from Salmone lla typhimurium (456 aa), fasta scores; opt: 169 z-score: 1 82.8 E(): 0.0055, 24.2% identity in 355 aa overlap"
/codon_start=1
/transl_table=11
/label=SC10A5.05
/protein_id="CAA16437.1"
/db_xref="GI:2815314"
/db_xref="SPTREMBL:O54098"
/translation="MSKLTDVPKRILIGRALRSDRLGETLLPKRIALPVFASDPLSSV
AYAPGEVLLVLSIAGVSAYHFSPWIAVAVVVLMFTVVASYRQNVHAYPSGGGDYEVAT
TNLGPKAGLTVASALLVDYVLTVAVSISSGIENLGSAIPFVVEHKVLCAVAVILLLTL
MNLRGVRESGTLFAIPTYVFVAGVFIMIVWGAFRGLVLDDTMRAPTADYEIKPEHGGL
AGFALIFLLLRAFSSGCAALTGVEAISNGVPAFRKPKSKNAGNTLAMMGLLAVTMFCG
IIALAAATDVRMSENPATDLFHNGVAVGADYVQHPVISQVAEAVFGEGSFLFIVLAAA
TALVLFLAANTAYNGFPLLGSILAQDRYLPRQLHTRGDRLAFSNGIVLLAGAAMLLVV
VYGADSTRLIQLYIVGVFVSFTLSQIGMVRHWNRNLAGERDQSKRRHMMRSRAINAFG
AFFTGLVLVVVLATKFTHGAWVALLGMCIFFATMTAIRKHYDRVAEEIAAPEDPEEAQ
SDDMVRPSRVHSVVLISKIHRPTLRALAYAKLMRSDSLEALSVNVDPAETKALREEWE
RRGIAVPLKVLDSPYREITRPVIEYVKSLRKESPRDAVSVIIPEYVVGHWYEHLLHNQ
SALRLKGRLLFTPGVMVTSVPYQLESSEAARRRARKRQDWSAPGAVRRGPAHHHQDRD
RTKDSSSST"

RBS 5812. .5815
/note="possible RBS upstream of SC10A5.06"

gene 5826. .7202
/gene="SC10A5.06"

CDS 5826. .7202
/gene="SC10A5.06"
/note="SC10A5.06, unknown, len: 458; similar to M. tubercul osis hypothetical protein TR:O07191 (EMBL:Z96072) MTCY05A6. 10C (405 aa), fasta scores; opt: 455 z-score: 878.5 E(): 0, 38.4% identity in 417 aa overlap, and weakly simialar to s everal putative RNA methyltransferases eg. YGCA_ECOLI P5513 5 hypothetical rna methyltransferase (433

Query Match 2.8%; Score 38.2; DB 1; Length 38734;
Best Local Similarity 48.0%; Pred. No. 1e+02;
Matches 109; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy   1116 ccgcgccatggtaccgatcggcacttatgagcgcgtaatgccgttggacatcctgcctac 1175
          | ||| |||  |  || | | | |        |||   |  || ||| | || | | ||
Db   4519 CTGCGGCATCATCGCGCTGGCCGCCGCGACCGACGTGCGGATGTCGGAGAACCCGGCCAC 4578

Qy   1176 cttgcttttgcgcgatttaatcgtcggcgataccgacagcgcgcaggctttgggttgctt 1235
          |   || || | | |     |||  | ||   | |||  ||  |||     ||    ||
Db   4579 CGACCTCTTCCACAACGGCGTCGCGGTCGGCGCGGACTACGTCCAGCACCCGGTGATCTC 4638

Qy   1236 ggaattggacgaagaagacctcgctttgtgcagcttcgtctgcccgggcaaatacgaata 1295
          | |  | | ||| |  | | |||    | ||||||||| | | |   |       || |
Db   4639 GCAGGTCGCCGAGGCGGTCTTCGGCGAGGGCAGCTTCCTGTTCATCGTGCTGGCCGCAGC 4698

Qy   1296 cggcccgctgttgcgcaaagtgctggaaaccattgagaaggaaggct 1342
          |  | ||||| | | |     |    |  | ||  | | |  | ||||
Db   4699 CACCGCGCTGGTCCTCTTCCTCGCCGCCAACACCGCGTACAACGGCT 4745
```

RESULT 35
AB015974/c
LOCUS AB015974 1641 bp DNA linear BCT 28-AUG-1999
DEFINITION Pseudomonas tolaasii glpK gene for glycerol kinase, complete cds.
ACCESSION AB015974
VERSION AB015974.1 GI:3358090
KEYWORDS glpK; glycerol kinase.
SOURCE Pseudomonas tolaasii (strain:PT814) DNA, clone_lib:pHHM105.
ORGANISM Pseudomonas tolaasii
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (sites)
AUTHORS Murata,H.
TITLE Molecular cloning of glpFKRD of a mushroom pathogenic bacterium, P seudomonas tolaasii strain PT814: induction of an avirulent mutant carrying single mini-Tn5km1 insertion in glpD
JOURNAL Mycoscience 40, 351-356 (1999)
REFERENCE 2 (bases 1 to 1641)
AUTHORS Murata,H.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1998) Hitoshi Murata, Forestry and Forest Products Research Institute, Bio-Resource Development; Matsunosato-1, Kukizaki-machi, Ibaraki 305-8687, Japan (E-mail:murmur@ffpri.affrc.go.jp, Tel:81-298-73-3211(ex.458), Fax:81-298-73-3795)
COMMENT On Jul 30, 1998 this sequence version replaced gi:3299803.
FEATURES Location/Qualifiers
source 1. .1641
/organism="Pseudomonas tolaasii"
/strain="PT814"
/db_xref="taxon:29442"
/clone_lib="pHHM105"

RBS 71. .74

gene 83. .1594
/gene="glpK"

CDS 83. .1594
/gene="glpK"
/codon_start=1
/transl_table=11
/product="glycerol kinase"
/protein_id="BAA31995.1"
/db_xref="GI:3358091"
/translation="MTDIQNKNYIIALDQGTTSSRAIIFDRDANVVCTAQREFTQHYP
QAGWVEHDPMEIFATQSAVMVEALAQAGLHHDQVAAIGITNQRETTVVWDKVTGRPIY
NAIVWQCRRSTEIQICQQLKRDGHEQYINDTTGLVTDPYFSGTKLKWILDNVEGSRER
ARNGELLFGTIDSWLIWKFTGGKTHVTDYTNASRTMLFNIHTLEWDAKMLEILDVPRE
MLPEVKSSSEIYGRTKVASPSAAIAGDQQAALFGQMCVRGGQAKNTYGTGCFLLMNTG
DKAVKSKHGMLTTIACGPRGEVAYALEGAVFNGGSTVQWLRDELKIIADATDTEYFAG
KVKDSNGVYLVPAFTGLGAPYWDPYARGALFGLTRGVRVDHIIRAALESIAYQTRDVL
DAMQQDSGERLKALRVDGGAVANNFLMQFQADILGTQVERPQMRETTALGAAYLAGLA
CGFWGSLEELRGKAVIEREFEPQLDEAAKEKLYAGWQKAVSRTRDWEPHEGAE"
BASE COUNT 346 a 541 c 492 g 262 t
ORIGIN

Query Match 2.8%; Score 38; DB 1; Length 1641;
Best Local Similarity 46.9%; Pred. No. 91;
Matches 119; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy   1007 tcggctgggttgcgccgcagccggacaaatactccatcacgcgcaccactctcggccatt 1066
          ||||    |  ||| | | ||| |||||   |   ||    ||  |  |    | || |
Db    423 TCGGTGCTGCGGCGGCACTGCCAGACAATGGCGTTATAGATCGGGCGGCCGGTGACCTTG 364

Qy   1067 tcctaaaaaacaaactcttcaagttcacgacagccgtcaacggcggcgaccgcgccatgg 1126
          ||| | |  ||      |||| | |       |  | | | ||||||||||    |||||
```

```
Db    363 TCCCACACCACCGTGGTTTCACGCTGGTTGGTGATGCCGATGGCGGCGACCTGATCATGG 304

Qy   1127 taccgatcggcacttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgc 1186
          | | |  ||||       |||||  |    |       |   |||  |  |     ||| |
Db    303 TGCAGGCCGGCCTGGGCCAGCGCTTCGACCATCACCGCGCTCTGGGTGGCGAAAATTTCC 244

Qy   1187 gcgatttaatcgtcggcgataccgacagcgcgcaggctttgggttgcttggaattggacg 1246
             | ||  ||| |   |||| |     |   |    || ||   ||     ||| ||
Db    243 ATCGGGTCATGCTCGACCCAACCGGCTTGCGGGTAGTGCTGAGTGAATTCACGCTGGGCG 184

Qy   1247 aagaagacctcgct 1260
             | |||||  || |
Db    183 GTGCAGACCACGTT 170


RESULT  36
SME591791/c
LOCUS       SME591791              340900 bp    DNA     linear   BCT 16-AUG-2001
DEFINITION  Sinorhizobium meliloti 1021 complete chromosome; segment 10/12.
ACCESSION   AL591791 AL591688
VERSION     AL591791.1  GI:15075538
KEYWORDS    .
SOURCE      Sinorhizobium meliloti.
  ORGANISM  Sinorhizobium meliloti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Sinorhizobium.
REFERENCE   1  (bases 1 to 340900)
  AUTHORS   Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
            Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
            Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
            Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
            Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
  TITLE     From the Cover: Analysis of the chromosome sequence of the legume
            symbiont Sinorhizobium meliloti strain 1021
  JOURNAL   Proceedings of the National Academy of Sciences of the United
            States of America. 98 (17), 9877-9882 (2001)
   PUBMED   11481430
REFERENCE   2  (bases 1 to 340900)
  AUTHORS   Gouzy,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
            EU Consortium
COMMENT     MELILO EU Consortium:
            Laboratoire de Biologie Moleculaire des Relations
            Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
            France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
            Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
            France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
            Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
            D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
            Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
            B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
            des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
            B-5030 Gembloux, Belgium.  E-mail:Jerome.Gouzy@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/meliloti.html.
FEATURES             Location/Qualifiers
     source          1. .340900
                     /organism="Sinorhizobium meliloti"
                     /strain="1021"
                     /db_xref="taxon:382"
     gene            132. .518
                     /gene="SMc01981"
     CDS             132. .518
                     /gene="SMc01981"
                     /function="small molecule metabolism; energy transfer;
                     electron transport"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PUTATIVE CYTOCHROME C TRANSMEMBRANE PROTEIN"
                     /protein_id="CAC47094.1"
                     /db_xref="GI:15075539"
                     /translation="MRIALAALSFSALSLCPVAGLAQEGDAEAGATVFKKCATCHVID
                     KDQNKVGPSLQGVIGRTAGTHADFKYSQAMIDAGKGGLVWDDATLAEYLRNPRAKVKG
                     TKMVFPGLKKDEEIANVIAYLKQHPK"
     gene            656. .1600
                     /gene="coxM OR SMc01982"
     CDS             656. .1600
                     /gene="coxM OR SMc01982"
                     /EC_number="1.9.3.1"
                     /function="small molecule metabolism; energy transfer;
                     electron transport"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE
                     POLYPEPTIDE II TRANSMEMBRANE PROTEIN"
                     /protein_id="CAC47095.1"
                     /db_xref="GI:15075540"
                     /translation="MAVVVILVLLAVGSVLFHLLSPWWWTPIASNWNYIDNTITITFW
                     ITGIAFTAVVLFMAYCVLRFRHRPGNTAAYEPENRRLEGWLATGTTFGVAAMLAPGLF
                     VWNQFVTVPQDASEVEVIGQQWLWSFRLPGADGKLGTTETRDIAPENTLGVNRDDAAG
                     QDDIIIEGGELHLPVGKPVKMLLRSVDVLHDFYVPEFRAKMDMVPGMITYFWLTPTRT
                     GTFEILCAELCGVGHPQMRGTVVVDTEEDYQAWLAEQQTFSQLSASSETRAVPEKVCS
                     GFPSGIATEQGTGASALFKEERECFGPAATTVAASAAQ"
     repeat_region   complement(1457. .1513)
                     /note="Sm-5 OR SMc04646
                     REPEAT SM-5
                     predicted by Homology"
                     /evidence=not_experimental
     gene            1634. .3415
                     /gene="coxN OR SMc01983"
     CDS             1634. .3415
                     /gene="coxN OR SMc01983"
                     /EC_number="1.9.3.1"
                     /function="small molecule metabolism; energy transfer;
                     electron transport"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE
                     POLYPEPTIDE I TRANSMEMBRANE PROTEIN"
                     /protein_id="CAC47096.1"
                     /db_xref="GI:15075541"
                     /translation="MMVDVRSGIGEALPPPEVEDVELYHPHSWWTRYVFSQDAKIIAI
                     QYSMTAIAIGMVALVLSWLIRLQLGFPGTFELIDAERYYQFITMHGMIMVIYLLTALF
                     LGGFGNYLIPLMVGARDMVFPYANMLSYWIYLLAVIVLAASFFTPGGPTGAGWTLYPP
                     QAVLSGTPGGRDWGIIMMLSSLIIFVIGFTMGGLNYVVTVLQGRARGMTLMRLPLTVW
                     GIFTATVMALLAFPALFVACVMMLFDRLLGTSFFMPAIVEMGEQLQYGGGSPILFQHL
                     FWFFGHPEVYIVALPAFGIVSDLISTHARKNIFGYRMMVWAIVIIGGLSFIVWAHHMY
                     VSGMNPYFGFFFATTTLIIAVPTAIKVYNWVLTLWRGNIHLTLPMLFALAFIVTFVNG
                     GLTGLFLGNVVVDVPLSDTMFVVAHFHMVMGVAPIMVIFGAIYHWYPKITGRMLNEAM
                     GQIHFWVTFIGAYAIFFPMHYLGLIGVPRRYHELGEASFVTTSIAELNAFISVMALLV
                     GAAQIVFLFNLAWSLRHGREAGGNPWRATTLEWQTPETPPPHGNWGNELPVVYRWAYD
                     YSVPGAPEDFIPQNQPTPGRLSHETVS"
     gene            3412. .4110
                     /gene="coxO OR SMc01984"
     CDS             3412. .4110
                     /gene="coxO OR SMc01984"
                     /EC_number="1.9.3.-"
                     /function="small molecule metabolism; energy transfer;
```

```
                electron transport"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE CYTOCHROME-C OXIDASE TRANSMEMBRANE
                PROTEIN"
                /protein_id="CAC47097.1"
                /db_xref="GI:15075542"
                /translation="MSIVAFFLAAIAAIIAWWLAGQRLTSRPWLEVGHFHDRRGATRL
                PPAKIGLGVFLAVVGALFSLAISAYFMRMASSDWGALPLPGLLWLNTGILAAGSITLH
                WTKVEAERRNDEAARIGLLAGLALGLAFLAGQLFAWRALSDAGYFLAGNPANSFFYLL
                TGMHGLHIIGGLFALGRVTAHASQTPLGNRTRLSIELCAIYWHFMLIVWLVLFALFAG
                WASGVIEFCRQLLT"
gene            4123. .4845
                /gene="coxP OR SMc01985"
CDS             4123. .4845
                /gene="coxP OR SMc01985"
                /EC_number="1.9.3.-"
                /function="small molecule metabolism; energy transfer;
                electron transport"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE CYTOCHROME-C OXIDASE PROTEIN"
                /protein_id="CAC47098.1"
                /db_xref="GI:15075543"
                /translation="MSAPMKKPAAETLPRPAGLAGLASDWASDQRTFKDVSWGKAMMW
                IFLLSDTFVFGCFLLAYMSARMSTSVPWPNPSEVFALEIGGTHMPLILIAIMTFVLIS
                SSGTMAMAVNYGYRRDRRKTAALMLLTALFGAAFVGMQAFEWSKLIAEGVRPWGNPWG
                AAQFGSTFFMITGFHGTHVTFGVIFLLIVARKVWRGDFETERRGFFTSRKGRYEIVEI
                TGLYWHFVDLVWVFIFAFFYLW"
gene            4857. .5216
                /gene="SMc01986"
CDS             4857. .5216
                /gene="SMc01986"
                /function="miscellaneous; hypothetical/partial homology"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                /protein_id="CAC47099.1"
                /db_xref="GI:15075544"
                /translation="MAHAETHPAQHAAAHTEHQQHAIKLYLLVWGLLFVLSALSYLVD
                YFGLQGYLRWSLILIFMMLKAGLIVAVFMHMAWERLALIYAIILPPLLVLVFVALMVS
                ESNYVLFTRLTFFGGGE"
gene            5477. .8038
                /gene="SMc01987"
CDS             5477. .8038
                /gene="SMc01987"
                /function="small molecule metabolism"
                /note="Product confidence : putative
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11

Query Match             2.8%;  Score 37.8;  DB 1;  Length 340900;
  Best Local Similarity   54.7%;  Pred. No. 1.5e+02;
  Matches   75;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    403 ccgttcagcaaaatccctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatg 462
          ||| ||   |   || | | | |    | || |       | | | |  | ||   | ||
Db 186372 CCGGTCTCGATCCTCACCGGCTTTCTCGGCGCGGGCAAGACGACCCTTCTGAACCGGGTG 186313

Qy    463 gacaccaatccgctggctgccgaccctacggtcatcatcaaagaagccgccgaagacttc 522
            ||   ||||  |  |||||| |  | |||||||||| ||| || ||| | || |
Db 186312 CTCAAGGATCCGGCGCTTGCCGACACGGCAGTCATCATCAACGAATTCGGCGATGTCTCC 186253

Qy    523 aaacgcggcctgttggt 539
          |    |  |||  ||||
Db 186252 ATCGACCACCTCCTGGT 186236


RESULT  37
AY011741
LOCUS       AY011741                760 bp    DNA     linear   MAM 14-FEB-2001
DEFINITION  Canis familiaris EDG1 (EDG1) gene, partial cds.
ACCESSION   AY011741
VERSION     AY011741.1  GI:12829947
KEYWORDS    .
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Murphy,W.J., Eizirik,E., Johnson,W.E., Zhang,Y.P., Ryder,O.A. and
            O'Brien,S.J.
  TITLE     Molecular phylogenetics and the origins of placental mammals
  JOURNAL   Nature 409 (6820), 614-618 (2001)
  MEDLINE   21082082
REFERENCE   2  (bases 1 to 760)
  AUTHORS   Murphy,W.J., Eizirik,E., Johnson,W.E., Zhang,Y.P., Ryder,O.A. and
            O'Brien,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2000) Laboratory of Genomic Diversity, National
            Cancer Institute-FCRDC, P.O. Box B, Bldg. 560, Rm. 11-10,
            Frederick, MD 21702, USA
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Canis familiaris"
                     /db_xref="taxon:9615"
     mRNA            <1. .>760
                     /gene="EDG1"
                     /product="EDG1"
     gene            <1. .>760
                     /gene="EDG1"
     CDS             <1. .>760
                     /gene="EDG1"
                     /codon_start=3
                     /product="EDG1"
                     /protein_id="AAK02009.1"
                     /db_xref="GI:12829948"
                     /translation="VFILICCFIILENIFVLLTIWKTKKFHRPMYYFIGNLALSDLLA
                     GVAYTANLLLSGATTYKLTPAQWFLREGSMFVALSASVFSLLAIAIERYITMLKMKLH
                     NGSNSFRSFLLISACWVISLVLGGLPIMGWNCIGALASCSTVLPLYHKHYILFCTTVF
                     TLLLLAIVILYCRIYSLVRTRSRRLTFRKNISKASRSSEKSLALLKTVIIVLSVFIAC
                     WAPLFILLLLDVGCKVKTCHILFRAEYFLVLAVL"
BASE COUNT      127 a    262 c    188 g    183 t
ORIGIN


  Query Match             2.8%;  Score 37.6;  DB 4;  Length 760;
  Best Local Similarity   48.6%;  Pred. No. 1.1e+02;
  Matches  103;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy   1010 gctgggttgcgccgcagccggacaaatactccatcacgcgcaccactctcggccatttcc 1069
          |||   |  ||  || |  |   |||  |  || |  | ||| ||   || ||   ||||
Db    277 GCTACATCACGATGCTGAAGATGAAACTCCACAACGGGAGCAACAGCTTCCGCTCCTTCC 336

Qy   1070 taaaaaacaaactcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtac 1129
```

```
Db    337 TGCTCATCAGCGCCTGCTGGGTCATCTCCCTGGTCCTGGGCGGCCTGCCCATCATGGGCT 396

Qy   1130 cgatcggcacttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcg 1189

Db    397 GGAACTGCATCGGCGCGCTGGCCAGCTGCTCCACCGTGCTGCCGCTCTACCACAAGCACT 456

Qy   1190 atttaatcgtcggcgataccgacagcgcgcag 1221

Db    457 ATATCCTCTTCTGCACCACCGTCTTCACGCTG 488
```

RESULT 38
KPU81143
LOCUS KPU81143 1731 bp DNA linear BCT 31-MAR-1998
DEFINITION Klebsiella pneumoniae heat shock protein (HSP60Kp) gene, complete cds.
ACCESSION U81143
VERSION U81143.1 GI:3002493
KEYWORDS .
SOURCE Klebsiella pneumoniae.
ORGANISM Klebsiella pneumoniae
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Klebsiella.
REFERENCE 1 (bases 1 to 1731)
AUTHORS Cancino-Diaz,M.E., Curiel-Quezada,E., Cancino-Diaz,J.C., Dominguez-Lopez,M.L. and Garcia-Latorre,E.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-1996) Inmunologia, Escuela Nacional de Ciencias Biologicas. Instituto Politecnico Nacional, Prolongacion de Carpio y Plan de Ayala s/n, Mexico City 11340, Mexico

```
FEATURES             Location/Qualifiers
     source          1. .1731
                     /organism="Klebsiella pneumoniae"
                     /db_xref="taxon:573"
     gene            88. .1731
                     /gene="HSP60Kp"
     CDS             88. .1731
                     /gene="HSP60Kp"
                     /note="groEL-like"
                     /codon_start=1
                     /transl_table=11
                     /product="heat shock protein"
                     /protein_id="AAC08714.1"
                     /db_xref="GI:3002494"
                     /translation="MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSF
                     GAPTITKDGVSVAREIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIVNE
                     GLKAVAAGMNPMDLKRGIDKAVLAAVEELKALSVPCSDSKAIAQVGTISANSDETVGK
                     LIAEVDKVGKEGVITVEDGTGLEDELDVVEGMQFDRGYLSPYFINKPDTGAVELESPF
                     ILLADKKISNIREMLPVLEAVAKAGKPLVIIAEDVEGEALATLVVNTMRGIVKVAAVK
                     APGFGDRRKAMLQDIATLTGGTVISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDG
                     VGEESAIQGRVAQIRKQIEEATSDYDREKLQERVAKVAGGVAVIKVGAATEVEMKEKK
                     ARVDDALHATRAAVEEGVVAGGGVALVRVAAKLAGLTGQNEDQNVGIKVALRAMEAPL
                     RQIVSNAGEEPSVVANNVKAGDGNYGYNAATEEYGNMIDFGILDPTKVTRSALQYAAS
                     VAGLMITTECMVTDLPKSDAPDLGAAGGMGGMGGMGMM"
BASE COUNT 457 a 437 c 513 g 324 t
ORIGIN
```

```
  Query Match          2.8%;  Score 37.2;  DB 1;  Length 1731;
  Best Local Similarity  45.5%;  Pred. No. 1.5e+02;
  Matches  132;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy    267 agtcgtgattgccgttgaaggcaacgacgaaatcgagttcgaacgctacgtacctgaagc 326

Db    600 AGGCGTGATCACCGTTGAAGACGGTACCGGTCTGGAAGACGAACTGGACGTAGTTGAAGG 659

Qy    327 gctggcaaaattgagcagcgaaaaagtgcgccgcaacctgattcaatcaggcttatggac 386

Db    660 TATGCAGTTCGACCGCGGCTACCTGTCTCCGTACTTCATCAACAAGCCGGACACCGGTGC 719

Qy    387 tgcgcttcgcacccgtccgttcagcaaaatccctgccgtagatgccgagccgttcgccat 446
```

```
Db    720 CGTTGAGCTGGAAAGCCCGTTCATCCTGCTGGCTGACAAAAAAATCTCCAACATCCGCGA 779

Qy    447 cttcgtcaatgcgatggacaccaatccgctggctgccgaccctacggtcatcatcaaaga 506

Db    780 AATGCTGCCGGTTCTGGAAGCCGTAGCGAAAGCTGGCAAACCGCTGGTCATCATCGCTGA 839

Qy    507 agccgccgaagacttcaaacgcggcctgttggtattgagccgcctgaccg 556

Db    840 AGACGTTGAAGGCGAAGCGCTGGCGACCCTGGTGGTTAACACCATGCGCG 889
```

RESULT 39
AF190427/c
LOCUS AF190427 4130 bp DNA linear PLN 25-OCT-2000
DEFINITION Colletotrichum lindemuthianum putative GAL4-like transcriptional activator (CLTA1) gene, complete cds.
ACCESSION AF190427
VERSION AF190427.1 GI:10998377
KEYWORDS .
SOURCE Colletotrichum lindemuthianum.
ORGANISM Colletotrichum lindemuthianum
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae; mitosporic Phyllachoraceae; Colletotrichum.
REFERENCE 1 (bases 1 to 4130)
AUTHORS Dufresne,M., Perfect,S., Pellier,A.L., Bailey,J.A. and Langin,T.
TITLE A GAL4-like protein is involved in the switch between biotrophic and necrotrophic phases of the infection process of Colletotrichum lindemuthianum on common bean
JOURNAL Plant Cell 12 (9), 1579-1590 (2000)
PUBMED 11006333
REFERENCE 2 (bases 1 to 4130)
AUTHORS Dufresne,M., Perfect,S., Pellier,A.L., Bailey,J.A. and Langin,T.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-1999) Sainsbury Laboratory, John Innes Center, Colney Lane, Norwich NR4 7UH, United Kingdom

```
FEATURES             Location/Qualifiers
     source          1. .4130
                     /organism="Colletotrichum lindemuthianum"
                     /strain="UPS9"
                     /db_xref="taxon:5458"
     mRNA            join(801. .1591,1654. .1673,1750. .1850,1902. .2129,
                     2180. .2794,2844. .4130)
                     /gene="CLTA1"
                     /product="putative GAL4-like transcriptional activator"
     gene            801. .4130
                     /gene="CLTA1"
     CDS             join(1522. .1591,1654. .1673,1750. .1850,1902. .2129,
                     2180. .2794,2844. .4050)
                     /gene="CLTA1"
                     /note="belongs to the zinc cluster protein family of
                     fungal transcriptional activators"
                     /codon_start=1
                     /product="putative GAL4-like transcriptional activator"
                     /protein_id="AAG25917.1"
                     /db_xref="GI:10998378"
                     /translation="MPRSSFSDNPLLRVSRPVSACSRCRAAKVKCDGKLPACTACEKA
                     GRENECSSANDQFARGKERSYVAALELRIEKLEKRLAFARSCKASVNLNDVDEAAAPA
                     PADRKDSLANIRAAIHRKAARKREDADVNTLISDMAYLTVSTTTRDFDPTMSNMTFAR
                     MVLAAATNDDVPDPAEDSFPTRAEALATIQYYTANIFTLFPFFSETHLLTVLDVVFQQ
                     DERAKEGLKDTDFWLLYMVLAVGYTAQSKAHDDNFYRRGLDFVARALPFADNALTPGY
                     PAQIQSLLLFTQYAMLDPSHFDSWYLIGFASRAVVDLGFHQDPPRLQVPDKVAFDLRR
                     KMFYCVYALDRTISMIRTRSFSFTDDSVNVEFPSPSGPSTTLITGPQSADPALLLFQL
                     RRIQSEWYQTLFQANSGPLPDAASFIWQMCLEMREWNESLPDNLPSGIRELLGLELRY
                     SYVYCLAPSDRSPHLTDYARKLIFEHAIAYMNTAHGVAHGTSTAFYTSLDALKVYFVA
                     MQFFAVFNAARELLISDAENHSSHHSPRKRTAPAFPTPSGGWRGQCRQSLRCLQQVVE
                     TLSRFSERWDIASQLKTTFETISSELFSWLQVRRQMRDQQPPQTPQQQQHPHQHQHQQ
                     HQQQQHQPQQHQQQQHQQQQHQLQHQQFSPGPPPPHPHHHQSHASYPTPQPPQPHPHH
                     QPHPSYPPPQHQPHPHSQVIMGHPAMQIPGMMPQPGQPMVQYRWVGDGGPMMQVPPPG
                     QGGPPR"
BASE COUNT 894 a 1274 c 1052 g 910 t
ORIGIN
```

```
  Query Match             2.8%;  Score 37.2;  DB 8;  Length 4130;
  Best Local Similarity   44.4%;  Pred. No. 1.6e+02;
  Matches  150;  Conservative    0;  Mismatches  188;  Indels    0;  Gaps    0;

Qy     251 agcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatcgagttcgaac 310
           |||   |  | |||    |||    |    |  |  ||  ||||| ||   ||
Db    1949 AGCCTCTTTTCCAGCTTCTCGATTCTCAGCTCAAGAGCCGCGACGTAACTAAGAGGTTTT 1890

Qy     311 gctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgcaacctgattc 370
             |||||  |    || || || |     ||  ||||     || |     |  |
Db    1889 AGCACGTAGTTCGGACGGTGCCACATCGAGGCGACGAACCTTCGCTCTTTTCCGCGGGCG 1830

Qy     371 aatcaggcttatggactgcgcttcgcacccgtccgttcagcaaaatccctgccgtagatg 430
           ||   | | | || |||      | ||  |   ||  | ||    || | | |      |
Db    1829 AACTGGTCGTTTGCACTAGAACACTCATTCTCGCGGCCGGCCTTCTCACAGGCAGTACAG 1770

Qy     431 ccgagccgttcgccatcttcgtcaatgcgatggacaccaatccgctggctgccgacccta 490
           |  |  ||| ||||||    | |    ||   | ||  ||| ||  |||||| |
Db    1769 GCCGGAAGTTTGCCATCACACTAAGGCGGAACAAGACGCATCAGCAAGCTGCCGTCAAAT 1710

Qy     491 cggtcatcatcaaagaagccgccgaagacttcaaacgcggcctgttggtattgagccgcc 550
           |  |||   |  | ||  | | ||| |  |  | ||   | || |||        ||
Db    1709 CCATCAGGCTGTATGAGCCGGGCGAGGTATATATACCTTGACTTTTGCGGCACGACCTGT 1650

Qy     551 tgaccgaacgtaaaatccatgtgtgtaaagcagcaggc 588
           |  |  |     | | |||| |   |||||||   ||
Db    1649 GGGGCTGATTAGCATTGCATGAGATGAAAGCAGTCAGC 1612


RESULT  40
I66494/c
LOCUS       I66494                  7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             2.8%;  Score 37.2;  DB 6;  Length 7218;
  Best Local Similarity   3.2%;  Pred. No. 1.6e+02;
  Matches   12;  Conservative  200;  Mismatches  158;  Indels    0;  Gaps    0;

Qy       7 aaaatcaaaaaaggtctaaatctgcccatcgcgggcagaccggagcaagtcatttatgac 66
           ::::  ::::::::   :::   :   :  : ::: :::  :::: :::  :   : ::
Db    1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356

Qy      67 ggcccggccattaccgaagtcgcgttgcttggcgaagaatatgtcggcatgcgcccctcg 126
           ::   ::  :  :  ::::  : :  :   :: :::::: : :  :: : : :      :
Db    1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296

Qy     127 atgaaaatcaaggaaggtgaagccgtcaaaaaaggccaagtgctgtttgaagacaaaaag 186
           : :::::  :::::::: ::::  :  :::::::::  ::: :  :   ::::: ::::::
Db    1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236

Qy     187 aatccgggcgtagtatttactgcgccggcttcaggcaaaatcgccgctattcaccgtggc 246
           ::   ::: : :: :   :  : :  ::    ::: ::::  :  :  :   :  : ::
Db    1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176

Qy     247 gaaaagcgcgtacttcagtcagtcgtgattgccgttgaaggcaacgacgaaatcgagttc 306
           :::::: : : :    ::  ::  : ::  :  :  ::::: :: :: ::::  :::
Db    1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116

Qy     307 gaacgctacgtacctgaagcgctggcaaaattgagcagcgaaaaagtgcgccgcaacctg 366
           ::: :  : : :   :::: :  :: ::::  ::: :: :::::::  :   |||||  ||
Db    1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTC 1056

Qy     367 attcaatcag 376
             || | | |
Db    1055 CCTCGACCTG 1046


RESULT  41
AL646079
LOCUS       AL646079              213050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
            segment 4/11.
ACCESSION   AL646079 AL646053
VERSION     AL646079.1  GI:17430956
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 213050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 213050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .213050
                     /organism="Ralstonia solanacearum"
                     /plasmid="megaplasmid"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     gene            100. .1632
                     /gene="RSp0487"
                     /note="RS00365"
     CDS             100. .1632
                     /gene="RSp0487"
                     /function="macromolecule metabolism; macromolecule
                     degradation; degradation of proteins, peptides,
                     glycopeptides"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD17638.1"
                     /db_xref="GI:17430957"
                     /translation="MDAPFSGRGAVIRRAVGRLLLVLVVGLAVLAVCIAFNTWRHDSR
```

```
                QLDVPPAAPLAVDGAAAAARLGEAIRARTISSFTDAQQNADQFRQLHAMLAARYPKAH
                AAMQREPVGDFSLLYTWKGSDPSLKPILLMAHQDVVPVAPGTEGDWTEPPFAGVVKDG
                MVWGRGAWDDKGNLIAQMEAAELLAASGFRPRRTIHFAFGADEEVGGERGAARIAALL
                KSRGEQLAFVIDEGLLITEGVVPGLTKPAALIGVTEKGFLSVALKLSATPGHSSMPPA
                PGESAIAMMSAALKHLDDQQLPAGIRGVTREMFETLAPEMRGLNRVALSNLWLFEPLV
                RKQLQASPSTNAVLQTTTALTIVQAGNKDNVLPGRAEATVNFRLLPGDSASSVIAHVE
                QAVRSAVPKGHFELAALPGVSEAAPVSPTQSASYQLIGRTVREVFPGTVVAPGLMVGA
                TDSRHMIGISDHVFRFSPVRARPEDLARFHGTNERISEANLVELIRFYHRLIHQAADA
                GA"
gene            complement(1751. .3148)
                /gene="czcS"
                /note="RSp0488; RS00366"
CDS             complement(1751. .3148)
                /gene="czcS"
                /EC_number="2.7.3.-"
                /function="cell processes; protection responses"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION
                HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE
                TRANSCRIPTION REGULATOR PROTEIN"
                /protein_id="CAD17639.1"
                /db_xref="GI:17430958"
                /translation="MMRWPASLAARLVLAFVLVAGVMLTGAGTYLYYALNAQLRATDD
                AMLLGKIVQLRHCAKDTQSIDDIVRDARRYTEVVSGHEGLILRVIGPDGRTLVEANPE
                GEPVAAPAAVPPDVLASESALHTWWSRTGIASRAIAAQARLGNHRDTVTVLVAQDGDS
                RAALMASYRRHIFLAVIAGVAAVALLSLWLVRSTLKGLARIGRQAASVVPSQLDTRLT
                LEGAPRELSDLLAALNAMLTRLQDGFARLSQFSADLAHDFRTPLANLIGQTEVTLAHP
                RPAGEYRAVLESSLEEYARLSRMIEDMLFLARADHARVALSVAPLDARHEIQVLVDYF
                EAVAADRDITIQVCGNGRVRADLTLLRRAINNLLDNALRHTPGGARIDVAVQVLPHST
                AIAVRNPGPGIPAASLPLLFDRFYRGDPARANSERSAGLGLAIVQTIMDLHGGTAQAS
                SPPGGLTEFRLIFPA"
gene            complement(3145. .3819)
                /gene="czcR"
                /note="RSp0489; RS00367"
CDS             complement(3145. .3819)
                /gene="czcR"
                /function="cell processes; protection responses"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
                CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
                /protein_id="CAD17640.1"
                /db_xref="GI:17430959"
                /translation="MRLLIVEDEPKAGDYLLKGLTESGFVADLARSGPDGLYQAMEHD
                YDLIVLDVMLPGMDGWQVIRELRRRKTTPVLFLTARDELSDRLKGLELGADDYLVKPF
                AFAELVARVHTILRRGPMRESEVLEIADLTIDTIKRRVTRAGRRIDLTAKEYALLHLL
                ARRTGEVLSRSLISSQVWDVNFDSNTNVVDVAIRRLRAKVDDPFEHKLIHTLRGMGYV
                LDTVAR"
gene            4089. .4394
                /gene="RSp0490"
                /note="RS00368"
CDS             4089. .4394
                /gene="RSp0490"
                /function="miscellaneous; hypothetical/global homology"
                /note="Product confidence : putative
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
                /protein_id="CAD17641.1"
                /db_xref="GI:17430960"
                /translation="MSVRRNCLLPLAAAAAFLACGAATAASVSEQALSPLQVHGARDP
                YTDGARGTRDPYTDGGAQAEATLQASVGPRDPYSDGGMSRSSAPRDRDPYTDGAQAA"
gene            4400. .5725
                /gene="czcC"
                /note="RSp0491; RS00369"
CDS             4400. .5725
                /gene="czcC"
                /function="cell processes; protection responses"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
                RESISTANCE PROTEIN"
                /protein_id="CAD17642.1"
                /db_xref="GI:17430961"
                /translation="MVTHPCIRAAVALSCALLVRSGLAAAPAGEAAAASDTAAPAAQS
                APASVPGAAHERMTLQQAIDRALATHPALAAAGYAVRASEGAVDQASRLRNPEAAFLT
                EDVRRDRATMTAQLNIPLELGGKRAARTRAAELARDAAGQDAAVLAAELRADVTRAFF
                ELSIAQERLRLAQASDALAAQAEAAAVRRVQAGKVSPVEQTKAAVARAGAAIELRDAQ
                ADLLVAQRALSAFWGGAVAVPEADGSAETLPAVPPEAHRVLAAATAAEPVALASTPRA
                LRARLEIEHRQALVNVERSKRVPDVTLSVGAKRDSGANANMAVVGIAVPLPLFDRNQG
                SLLEAQRLADKAAEDYRALRLEQAAALSQDVARLEAARQAVGALRQDVLPGAGRAYDA
                ARIGFEAGKFNFLDVLDAQRTLFQARAQYLAALSRAHQAAAGIDRILGR"
gene            5752. .6966
                /gene="czcB"
                /note="RSp0492; RS00370"
CDS             5752. .6966
                /gene="czcB"
                /function="cell processes; protection responses"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE COBALT-ZINC-CADMIUM RESISTANCE (CATION
                EFFLUX SYSTEM PROTEIN) TRANSMEMBRANE"
                /protein_id="CAD17643.1"
                /db_xref="GI:17430962"
                /translation="MTLPRQQRRAIALIVALGALAGAALWWGMGARQAAHAESEAEQP
                AAHAEHAQAPKAEAAEAPHAETIAMTAEAIANAAIGVETAGAAQLRTEVVLPGEIRLN
                DDRTAHVVPRVAGVAEQVAVELGQPVTRGQLLAVLSSPALSDQRSELQAAQQRLALAQ
                EAHAREKRLWEQGIAAAQDYQQARTALQEARIAADNARQKLAAIGAAPAGTALNRFEL
                RAPFDGVVVARHLSQGEAVQAEAAVFTIADLRTVWADFAVTAKDLEAVTTHATATVRA
                TATGTAVQGKVSYVGALLGEQTRSAPARVTLDNPKRAWRPGMFVSVSVESGRVPVPVA
                VRADAVQTVDEKPSVFVPVPAGFEVRPVKTGRSDGTWTEILDGLPAGTRYAAANSYVL
                KAEAGKSADHGH"
gene            6977. .10171
                /gene="czcA"
                /note="RSp0493; RS00371"
CDS             6977. .10171
                /gene="czcA"
                /function="cell processes; transport of small molecules"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage

  Query Match              2.8%;  Score 37.2;  DB 1;  Length 213050;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+02;
  Matches   93;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy   1136 gcacttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaa 1195
```

```
          || |  |  |   |||  ||  | || |  |||  |   ||  ||  ||  || |
Db  83911 GCTCGCAGAACGACGTGCTGGGGGTGCAGTTCCGCCTGTCCAAGCGGTTCGGCAAGAACT 83970

Qy   1196 tcgtcggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacc 1255
          |||||| ||| ||||| |   | |||     | |   |        |||  |||    | ||
Db  83971 TCGTCGACGACACCGAGATTCCCCAGCAGCGGCGCAACAACACGGTGGTGGAAACCGGCC 84030

Qy   1256 tcgctttgtgcagcttcgtctgcccgggcaaatacgaatacggcccgctgttgcgcaaag 1315
          |  |     ||  || | || |  | |   | | |||   | ||| |||| ||||| | |
Db  84031 TGACGGATCGCCACTACTTCGGGGCCGCGCAGTTCGACGGCAGCCTGCTGCTGCGCCACG 84090

Qy   1316 tgctgg 1321
             |||
Db  84091 GCGTGG 84096


RESULT  42
AE008180/c
LOCUS       AE008180              10029 bp    DNA     linear   BCT 18-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 circular chromosome, section 238
            of 254 of the complete sequence.
ACCESSION   AE008180 AE007869
VERSION     AE008180.1  GI:15157875
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (Cereon).
  ORGANISM  Agrobacterium tumefaciens str. C58 (Cereon)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 10029)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Complete Genome Sequence of Agrobacterium tumefaciens C58
            (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
            Disease in Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 10029)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
            Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1. .10029
                     /organism="Agrobacterium tumefaciens str. C58 (Cereon)"
                     /strain="C58"
                     /db_xref="taxon:181661"
     gene            157. .615
                     /gene="AGR_C_4831"
     CDS             157. .615
                     /gene="AGR_C_4831"
                     /note="Marr family regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4831p"
                     /protein_id="AAK88387.1"
                     /db_xref="GI:15157876"
                     /translation="MGTKREKDSLAFLLNSGARLLNNAFERRISEAGLGLTPGEARAL
                     LTVAVVDGSKQSDIAARLGIEPMTICAYLDKLQSLDLIERQQDPQDRRSKRIVLTKNS
                     TDMLEAVRQEIDELVRHATRGLNEDEVALFRSFLQTLRDNLQPEQPGQSC"
     gene            621. .1817
                     /gene="AGR_C_4834"
     CDS             621. .1817
                     /gene="AGR_C_4834"
                     /note="(AE006190) Bcr"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4834p"
                     /protein_id="AAK88388.1"
                     /db_xref="GI:15157877"
                     /translation="MPPEPIMSTKRTAILGALLTALAPISMAIYTPAMPELTRVFATN
                     ESAIKLSLSLYFAGFALAQLLAGPASDAFGRRRASLVFLSVFLGGGLLAIFAPTIEVL
                     LLARLIQGIGASVGMTVARAIVRDQFTGAEASSIMNLIGIMLAVGPALGPTIGGLSLA
                     AFGWKSVFFLMAGLGAAAIFSVVIFLRETTVPDKSLIRPRRLLSAYGNLLTEPRFLLA
                     ALVLGGSIGALYAQATMLTFILINTVGMTPTAFGIGMLMQTGAYFFGSIALRYIAPRL
                     GDRRSVMLGLCFSGTGGLLILLSVFLIPPSFFSIMGPVAVATVGIALLTPSMVTAALA
                     PFPHIAGSASAMMGFIQMGSGFAGGTAASLIGSPLTGFGIVIPVMEFTAVASYIGFRI
                     VSRSKT"
     gene            complement(1872. .3254)
                     /gene="AGR_C_4836"
     CDS             complement(1872. .3254)
                     /gene="AGR_C_4836"
                     /note="argininosuccinate synthase PA3525 {imported} -
                     Pseudomonas aeruginosa (strain PAO1)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4836p"
                     /protein_id="AAK88389.1"
                     /db_xref="GI:15157878"
                     /translation="MAGHHPSSCLPINCFVAYLTEKRFTHFGLHCKSAEIPRKWRCPD
                     AVIMQMDIIMALPKDVKKVVLAYSGGLDTSIILKWLQTELGAEVVTFTADLGQGEELE
                     PARKKAEMLGIKEIFVEDVREEFVRDFVFPMFRANAVYEGVYLLGTSIARPLISKHLI
                     EIAKKTGADAIAHGATGKGNDQVRFELSAYALNPDIKIIAPWRDWTFKSRTDLLNFAE
                     QHQIPVAKDKKGEAPFSVDANLLHSSSEGKVLEDPAVEAPEYVHMRTISPEAAPDKAT
                     VIKVGFRKGDACSINGVEMSPATLLATLNNYGRENGIGRLDLVENRFVGMKSRGVYET
                     PGGTILLAAHRAIESITLDRGAAHLKDDLMPRYAELIYYGFWFSPEREMLQALIDKSQ
                     EHVEGEVTLKLYKGNVMVTGRESEKSLYSDALVTFEDDHGAYDQKDAAGFIKLNALRL
                     RTLAKRNLNK"
     gene            3327. .4265
                     /gene="AGR_C_4837"
     CDS             3327. .4265
                     /gene="AGR_C_4837"
                     /note="probable oxidoreductase - Streptomyces coelicolor"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4837p"
                     /protein_id="AAK88390.1"
                     /db_xref="GI:15157879"
                     /translation="MKTVNEATQVQATLQTAGNSRNDAKEKPMTTIHPAFKEDNVAVI
                     TGAASGIGLAAARKFASFGMSVVLVDLDGEKLAAAHADILNVAKDGEAQIVAIPTDVS
                     KLDELEALERAVIQRFGRVHVLMNNAGIQPGSAIFGPQANWEKIIGVNLMGVVNGSRV
                     FGQGMITHGEPSLIINTGSKQGITTPPGDPAYNVSKSGVKTFTEALQHELRNIPNSAV
                     SAHLLIPGFVFTGLTAHGRTEKPQGAWTGEQTVDFMIESIGRGDFYILCPDGEVDRPT
                     DEKRILWAAQDIVQNRPPLSRWHTEYSAAFTSFLKN"
     gene            4338. .5045
                     /gene="AGR_C_4838"
     CDS             4338. .5045
                     /gene="AGR_C_4838"
                     /note="conserved hypothetical protein - Deinococcus
                     radiodurans  (strain R1)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4838p"
                     /protein_id="AAK88391.1"
                     /db_xref="GI:15157880"
                     /translation="MAETRNRVKQPHSPDFTGAAMDFVPSLPTLIAFTIAILLLAVTP
                     GPDMTLWISRSLREGRAAGFMTLVGTNIGITVHTMLVAFGVAALIVASPTAFMILKTG
                     GAAYLVWLAIQAIRKGSDFVMVKSTGEKAQASLKSALLNGIWVNLLNPKVIIFFMTFL
                     PQFVSATDPHVTGKLIFLGIWSIIVALPIGIGIVVTADILSAWLQRNRKVLRGLDYTI
                     AGVFSLFAVKIFFTQTR"
     gene            complement(5084. .5878)
                     /gene="AGR_C_4841"
     CDS             complement(5084. .5878)
                     /gene="AGR_C_4841"
                     /note="probable ATP-binding component of ABC transporter
                     PA3838 {imported} - Pseudomonas aeruginosa (strain PAO1)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4841p"
                     /protein_id="AAK88392.1"
                     /db_xref="GI:15157881"
                     /translation="MISLSDIQVVFGRGTPLQKQALAKIDLTIEDGSFVTVIGSNGAG
                     KSTLLGVLAGDVLPTAGKVMIGGADVTRKPTASRAGRVARVFQDPLAGSCGALTIEEN
                     LALAASRGERRGLSSALGRGRRDFFRERIASLNLGLENRLKDRMDLLSGGQRQAVSLV
                     MATLSGSDVLLLDEHTAALDPGMAEFVMELTRKVVSERKLTTLMVTHSMRQALDYGTR
                     TIMLHAGEIVLDVSGDSRKDLEVEDLIEMFRKIRGQTLDDDELLIG"
     gene            complement(5875. .6756)
                     /gene="AGR_C_4842"
     CDS             complement(5875. .6756)
```

```
                     /gene="AGR_C_4842"
                     /note="probable ABC transporter, permease protein VC1102
                     {imported} - Vibrio cholerae (group O1 strain N16961)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4842p"
                     /protein_id="AAK88393.1"
                     /db_xref="GI:15157882"
                     /translation="MSQIAFWGAVELGLVFAFVAIGVYLAFRVLDFPDLTVDGSFPLG
                     AAVAAVLIIAGLNAWVATAIAMVAGAAAGMVTATLNVRFKILNLLASILTMIALFSVN
                     LRVMGKPNVALLNQDTMVSPFYGLGLSEYLVRPLFVGVLVLIAVILVWRFLESDAGLA
                     MRATGANARMARAQGVKTGNQIYLGMALSNALVALGGALFAQTNGFADVTSGVGTIVV
                     GLAAVIIGETLFGARGILIALIGCVFGSIAYRLAIQLALSSDVLGLKASDLNFVTAVL
                     VAIALILPRLRRGGATS"
          gene       complement(6863. .7861)
                     /gene="AGR_C_4844"
          CDS        complement(6863. .7861)
                     /gene="AGR_C_4844"
                     /note="hypothetical protein PA3836 {imported} -
                     Pseudomonas aeruginosa (strain PAO1)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4844p"
                     /protein_id="AAK88394.1"
                     /db_xref="GI:15157883"
                     /translation="MFLERRKQGTNRRISMRAFILALAAASAIVLPAKAENVTVAVTA
                     IVEHPALDAARDGVKAALAEAGYKEGENLKFLYESAQGNPGTAAQIARQFVGDAPAVI
                     VPISTPSAQAVVAATRDIPVVFTAVSDPVGAQLIKSMEKPGGNVTGLSDMLPVGEHLA
                     LIKEISPNAKSIGFIYNSAEANSASTLALLKAEADKAGLKVIESVATKSAEVQGATRA
                     LVGKADVIYVPTDNTIVSAFEAVAGVASEAKIPLYAADTDSVARGAVAALGFNYFDVG
                     KQTGAVVVRILKGEKPGDIPATVAVGTDLVINKKVAEKVGVTFPESVTKRATKVID"
          gene       complement(7925. .9265)
                     /gene="AGR_C_4846"
          CDS        complement(7925. .9265)
                     /gene="AGR_C_4846"
                     /note="conserved hypothetical protein NMB1308 {imported} -
                     Neisseria meningitidis (group B strain MD58)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4846p"
                     /protein_id="AAK88395.1"
                     /db_xref="GI:15157884"
                     /translation="MREKCSAFDFLDMQTIFRLFDRPVVRCMERWDNVMSVMQAPAGL
                     AVKAPLIANSDLMSGKPSLIGLTREEMGEALAEIGVPQKQVKMRVSQLWNWLYVRGVS
                     DFDNMTNVAKELREKLKAAFTIARPEIVEEQISNDGTRKWLMRFPPRGAGRPVEIETV
                     YIPEEGRGTLCISSQVGCSLTCSFCHTGTQRLVRNLTAEEILSQLLLARDRLGDFPDG
                     STPVGAYVPSEGRKVSNIVMMGMGEPLYNFEHVKTALLIATDGDGLSLSKRRVTLSTS
                     GVVPEIFRTGDEIGVMLAISLHAVRDDLRDMLVPINKKYPLKELIEACRNYPGVSNAR
                     RITFEYVMLKDVNDSLEDAKMLVQLLKGVPAKINLIPFNPWPGTNYQCSEWAQIEKFA
                     DFINQAGYASPIRTPRGRDILAACGQLKSESERMRKTERLAFEAMMIANHGADD"
          gene       complement(9364. .9897)
                     /gene="AGR_C_4849"
          CDS        complement(9364. .9897)
                     /gene="AGR_C_4849"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_C_4849p"
                     /protein_id="AAK88396.1"
                     /db_xref="GI:15157885"
                     /translation="MEDNNPAGNSMFVRSVATAAAILLTSVGLASAQSPTRIQQFNAW
                     GAYSYKSGNSTVCYVLSIPTSKEPASVDHGDIFFIVSQRPGQNISYEPQAMVGYTLKQ
                     GSKVNVTIDNKNFVMFTKDKAAWVENAAEEPALVAAMKGGKSMTVKAVSGRGTATSYS
                     YSLSGISAAFKQIESCK"
BASE COUNT     2072 a   3048 c   2888 g   2021 t
ORIGIN


  Query Match             2.8%;  Score 37;  DB 1;  Length 10029;
  Best Local Similarity   48.8%;  Pred. No. 1.9e+02;
  Matches  100;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy    343 agcgaaaaagtgcgccgcaacctgattcaatcaggcttatggactgcgcttcgcacccgt 402
          | ||||||| |   |  |  | | | |||  | || |    | |  || | ||||| |
Db   8104 ATCGAAAAATTCGCCGACTTCATCAATCAGGCGGGTTATGCGTCGCCGATCCGCACGCCG 8045

Qy    403 ccgttcagcaaaatccctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatg 462
          |    | || | || |  ||||     | | ||  |    |    |  |       |
Db   8044 CGCGGCCGCGACATTCTCGCCGCCTGCGGCCAGTTGAAGTCGGAATCTGAACGCATGCGC 7985

Qy    463 gacaccaatccgctggctgccgaccctacggtcatcatcaaagaagccgccgaagacttc 522
           | ||| |    ||||||   ||  | | | | |||  |||  | | ||| || ||||
Db   7984 AAGACCGAAAGACTGGCTTTTGAGGCGATGATGATCGCCAACCACGGCGCAGATGACTGA 7925

Qy    523 aaacgcggcctgttggtattgagcc 547
           | | || || | |  |   | | ||
Db   7924 ATAAGCCGCGTTTGCGCCATAAACC 7900


RESULT  43
AE009214/c
LOCUS       AE009214               10359 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Agrobacterium tumefaciens strain C58 circular chromosome, section
            240 of 256 of the complete sequence.
ACCESSION   AE009214 AE008688
VERSION     AE009214.1  GI:17741176
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (U. Washington).
  ORGANISM  Agrobacterium tumefaciens str. C58 (U. Washington)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 10359)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     The Genome of the Natural Genetic Engineer Agrobacterium
            tumefaciens C58
  JOURNAL   Science 294 (5550), 2317-2323 (2001)
   PUBMED   11743193
REFERENCE   2  (bases 1 to 10359)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2001) Department of Microbiology, University of
            Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
            98195-7242, USA
FEATURES             Location/Qualifiers
     source          1. .10359
                     /organism="Agrobacterium tumefaciens str. C58 (U.
                     Washington)"
                     /strain="C58"
                     /db_xref="taxon:180835"
          gene       complement(84. .1037)
                     /gene="Atu2672"
          CDS        complement(84. .1037)
                     /gene="Atu2672"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
```

/protein_id="AAL43653.1"
/db_xref="GI:17741177"
/translation="MRAFILALAAASAIVLPAKAENVTVAVTAIVEHPALDAARDGVK
AALAEAGYKEGENLKFLYESAQGNPGTAAQIARQFVGDAPAVIVPISTPSAQAVVAAT
RDIPVVFTAVSDPVGAQLIKSMEKPGGNVTGLSDMLPVGEHLALIKEISPNAKSIGFI
YNSAEANSASTLALLKAEADKAGLKVIESVATKSAEVQGATRALVGKADVIYVPTDNT
IVSAFEAVAGVASEAKIPLYAADTDSVARGAVAALGFNYFDVGKQTGAVVVRILKGEK
PGDIPATVAVGTDLVINKKVAEKVGVTFPESVTKRATKVID"
gene complement(1146. .2318)
/gene="Atu2673"
CDS complement(1146. .2318)
/gene="Atu2673"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL43654.1"
/db_xref="GI:17741178"
/translation="MSGKPSLIGLTREEMGEALAEIGVPQKQVKMRVSQLWNWLYVRG
VSDFDNMTNVAKELREKLKAAFTIARPEIVEEQISNDGTRKWLMRFPPRGAGRPVEIE
TVYIPEEGRGTLCISSQVGCSLTCSFCHTGTQRLVRNLTAEEILSQLLLARDRLGDFP
DGSTPVGAYVPSEGRKVSNIVMMGMGEPLYNFEHVKTALLIATDGDGLSLSKRRVTLS
TSGVVPEIFRTGDEIGVMLAISLHAVRDDLRDMLVPINKKYPLKELIEACRNYPGVSN
ARRITFEYVMLKDVNDSLEDAKMLVQLLKGVPAKINLIPFNPWPGTNYQCSEWAQIEK
FADFINQAGYASPIRTPRGRDILAACGQLKSESERMRKTERLAFEAMMIANHGADD"
gene 2537. .3034
/gene="Atu2674"
CDS 2537. .3034
/gene="Atu2674"
/note="putative; ORF located using Glimmer"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL43655.1"
/db_xref="GI:17741179"
/translation="MAGRGPANHKSDKASGLLAALDLFESSRNARQRIGIGRRSAAAG
NRLYGHRLAALHGGDESRLFSGVFNPGSLVLGEHDEVLVVDRHIDLGALLQRIANHRL
RLVGDVLARTLRDDEEDIAVVNACRFLGGGDGKNVADGAVARFVRIGTPGIELLDARG
RLCRC"
gene complement(3386. .3766)
/gene="Atu2675"
CDS complement(3386. .3766)
/gene="Atu2675"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL43656.1"
/db_xref="GI:17741180"
/translation="MDDVKIGEILLPGETDKQRERENVVRAKFWPKLKRVMTKVPFAR
DAAAAYYCAIDRDTPLRAKGIILAALAYFIMPIDAVPDMLAVVGFTDDIAVITAALAM
IRAHIKMEHYEAADAMLKRQQEAG"
gene complement(3818. .4123)
/gene="Atu2676"
CDS complement(3818. .4123)
/gene="Atu2676"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="4a-hydroxytetrahydrobiopterin dehydratase"
/protein_id="AAL43657.1"
/db_xref="GI:17741181"
/translation="MKYPRLEPDAIRQALSKLEGWSLREDGAAIVRSFKFSSFAEAFG
FMAESALAAEKLNHHPEWSNVYSRVKVCLTTHDSQGVTERDFLLAEAMQKAAAGRGN"
gene 4183. .4662
/gene="Atu2677"
CDS 4183. .4662
/gene="Atu2677"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="protein tyrosine phosphatase"
/protein_id="AAL43658.1"
/db_xref="GI:17741182"
/translation="MKHITVLFVCMGNICRSPLAEGVLTDLADGEGVSDRITVDSAGT
GGWHTGDAPDPRSIAMARRHGIDISRQRARQVTRADFEAFDLILAMDENNLANLLQLS
PEKYRHKIHLFMDYAAGHPENVPDPYYGAEDGFLDVYNMLLAGCRSLLEKMELDRAS"
gene complement(4596. .5189)
/gene="ligT"
/note="Atu2678"
CDS complement(4596. .5189)
/gene="ligT"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="2'-5' RNA ligase"
/protein_id="AAL43659.1"
/db_xref="GI:17741183"
/translation="MPRLFTALEIPRNAAMSLSLLRGGLPGARWIDVENYHITLRFIG
DIDGRTADEVVDRLDRIERPEFQLNLTGIGSFGSKKPHSIWAGVSPSPEMHALQAEVE
RICQRIGLPPDPRKFMPHVTLARLRSSRVDDVVHYLSGRGNFRTSPFTVGRFVLMSSK
ESVGGGPYVVEEAFPLHEARSSSIFSSNDLHPAKSML"
gene complement(5258. .5908)
/gene="Atu2679"
CDS complement(5258. .5908)
/gene="Atu2679"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="esterase"
/protein_id="AAL43660.1"
/db_xref="GI:17741184"
/translation="MRFKAALFHFIVILAGAFSVASAVAQERTLQLVGLGDSLMAGYQ
LPPTDSYTAQLEAALKAKGINVAITNAGVSGDTSSGGLARAEWSVPDGTDGVILELGA
NDALRGIAPEQTEKNLDTIISGFQKRNIAVLLVGIMAPPNMGDDYAKRFNPIFPRLAE
KYNLPLYPFFLDGVVTDDALKLEDKMHPNTKGVAVMVEKSLPAVESFIKTIGAQKK"
gene 5999. .6709
/gene="Atu2680"
CDS 5999. .6709
/gene="Atu2680"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase
protein"
/protein_id="AAL43661.1"
/db_xref="GI:17741185"
/translation="MTKSIIKLKKADLTLGNAAASVHVLKNIDLSIDEGEAVGIVGPS
GSGKSTLLMVLAGLERLDSGEIVIADTQLHKLGEDALADFRGRNIGIVFQSFHLIANM
TALENVAVPLELANTPNPFEIAKRELVAVGLGERLNHYPGQLSGGEQQRVAIARALAP
SPAVLIADEPTGNLDTDTGKQIADLLFAKQAERGMTMVLVTHDPSLAARCSRQIKVRS
GEIEGDSARPQMARAVSA"
gene 6706. .9267
/gene="Atu2681"
CDS 6706. .9267
/gene="Atu2681"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL43662.1"
/db_xref="GI:17741186"
/translation="MIGSIFPSFTNIRNAARLARREIRGGLRGFYIFLACISLGTGAI
AAVNSVSRAVTSAIATEGQSILAGDVRFELRNREATAEERSYLDGLGTVAVSTGLRSM
ARLADGSEQTLTEVKAIDGAYPLFGTFVADPNRPLGELLAGSGDTYGAVAAPLLLERL
GIKVGDEILLGNAKLRLTGTVVDEPDALSDGFGFAPRLVVSRDALFASGLIQTGSLVE
HAYKIKLSDPAARATMTDAANKAFPQAGWSIRTSDRAAPSLTENVNRFSQFLTLVGLT

```
  Query Match             2.8%;  Score 37;  DB 1;  Length 10359;
  Best Local Similarity   48.8%;  Pred. No. 1.9e+02;
  Matches  100;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     343 agcgaaaaagtgcgccgcaacctgattcaatcaggcttatggactgcgcttcgcacccgt 402
           | ||||||| |   |  |  | | | |||  | || |    | |  || | ||||| |
Db    1325 ATCGAAAAATTCGCCGACTTCATCAATCAGGCGGGTTATGCGTCGCCGATCCGCACGCCG 1266

Qy     403 ccgttcagcaaaatccctgccgtagatgccgagccgttcgccatcttcgtcaatgcgatg 462
           |    | || | || |  ||||     | | ||  |    |    |  |       |
Db    1265 CGCGGCCGCGACATTCTCGCCGCCTGCGGCCAGTTGAAGTCGGAATCTGAACGCATGCGC 1206

Qy     463 gacaccaatccgctggctgccgaccctacggtcatcatcaaagaagccgccgaagacttc 522
           | ||| |    ||||||   ||  | | | | |||  |||  | | ||| || ||||
Db    1205 AAGACCGAAAGACTGGCTTTTGAGGCGATGATGATCGCCAACCACGGCGCAGATGACTGA 1146

Qy     523 aaacgcggcctgttggtattgagcc 547
           | | || || | |  |   | | ||
Db    1145 ATAAGCCGCGTTTGCGCCATAAACC 1121


RESULT  44
AE002070/c
LOCUS       AE002070               10514 bp    DNA     linear   BCT 22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 207 of 229 of the complete
            chromosome 1.
ACCESSION   AE002070 AE000513
VERSION     AE002070.1  GI:6460209
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcaceae;
            Deinococcus.
REFERENCE   1  (bases 1 to 10514)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 10514)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10514
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            93. .2717
                     /gene="DR2392"
     CDS             93. .2717
                     /gene="DR2392"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF11946.1"
                     /db_xref="GI:6460217"
                     /translation="MTVSDDTIGTNVQDGVRGVQRLTYIGVPLKGGTNVLRVLNDEIS
                     VSLVGSTAQIEVKPLALQADGSSPVRVQLRALDEFGKLTSQPTITVRTNLEPRLPDAN
                     PGEPGYQVRLTEGVGVLELQPQSAPTSLKVDVLQGEKVKTYRYEVTPDRSRVGIGVLS
                     ATLGLDGDFELQDDVTWQARGAYEGPLGAGKLYVAADKDGLPTDRNTLIRNTVYGDAS
                     TESVPLQGIDPVAFSYDHPQFRAEYRRTSLPIDVLPVGENLTALTAYSKTNPRVSGFV
                     ALVPEDRVVNRPLTPEGTRLLRLGDTGVSLGSETLELVTLERSTGKELQREKLVRNVD
                     YVLDPRSGVVTFNRAIDRLDAELNERVVYASYRLDNPLANRRTAYGAQIKTTGTNSSV
                     GVAAVYLDGRTTYGVKADYQRGPVKASGLLAYSGGVQASADLSTVLARNHDLNFRVRY
                     QQASYEGLAPFAPGLVVSGRYTGRLTPRLNAVADAEYRDMPSPRASDVENANHARGGN
                     VTARAVYHLDPFTVGAGVSYAFGDRSGLGLVGSVGYQREPISVDVVHTQPIIGDLDPT
                     TDISSRFRIRKDLTLGFNDKITWGVGQAAALTLDTMIGNVNYAVGYELPTASGAGNRA
                     RFGATTTLPLNDRTALNLRGNVLYDVGKGAFEGATGADINYKTDRLSATVGTDVSYRG
                     DTGFGVVLRGGVTGSVTDELTLNAGGTAEFGQGRRGERLELGYAYRGRSLSSLGYGRY
                     VAGSLAGGKPELSFGVSAAYRLPNWAVRGGVESRTQLDDRDSFTLQGWLGSSYYLTDR
                     LAVGAWARALTQPASQTTLFGYGVEGSVRPLPGTWLTAGYNFRGFEGFQSSSTYTRQG
                     PYLRLDLTLDETLGQPQR"
     gene            2787. .3701
                     /gene="DR2393"
     CDS             2787. .3701
                     /gene="DR2393"
                     /note="similar to GB:Pyro_h percent identity: 53.56;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF11939.1"
                     /db_xref="GI:6460210"
                     /translation="MTSHLRGILLLLLVTAIWGSTFAVVKELGALLAPPVLLAWRFSI
                     GALVLLPLAALRRTPAPTVTVTQADGTSLWSDGMVLGLWLIAGYGTQTIALQTTTANR
                     AAFFTALSVVLVPVWLTLVQRRRMPAVLWAALPLAVAGLALLSWEGGAWVSGDAWALA
                     CAVTYAGFILALEKLASRHAALPFTLAQVLSVALVAWGWALLSGAPLWPPQAAWAPLF
                     YLGVVATAGTTLLQTLGQRHVSAAEASLIYALEPVSAALFSFALIGERVGARGALGGA
                     LVVLATVLSSRAGETEPVARVLDSSQDD"
     gene            3894. .5702
                     /gene="DR2394"
     CDS             3894. .5702
                     /gene="DR2394"
                     /note="similar to PID:1177251 SP:P50864 PID:1644212
                     GB:AL009126 percent identity: 51.54; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="N-acetylmuramoyl-L-alanine amidase"
                     /protein_id="AAF11940.1"
                     /db_xref="GI:6460211"
                     /translation="MKLPAMLLSSALLLGSWAAAQTDPFLRAAPAQVKPTVRGSAAAP
                     PSAAPSFATPSAATQPLAQAKETTFGSPRISGTAEATKLVFDLPKGMTYTLVPTFGGL
                     RIDVGGVRVAPASSLRLAENLTEYRANNGQIIMGTPFPLSLSDGWRATEATIANGNRV
                     LIVELGSQLRGGAGNVGGRVLAAAPADPRAQASLQAPVPVTDTAASSALTSHLPPGDS
                     VKQQPATPPAPLAPLPGADTARPSALVGRMPGSLSSGATVTTPRIGKNPGVTRMVLDL
                     PPGAGYRIVPGTVGLRVELLGVGLGDANTNEQNTSAEVRAWRYEPSAEGVNLSIVTGS
                     PTTSRSGWRAQVLPPLNSERSRLVLDISPAMANLTPLSPRDKVLAAVPPMTASSGAAV
                     LSLRAASVRPRVIIDPGHGGKDPGATGAITEKDVTLDVARRVRDLLVAAGVDAVLTRE
                     TDRELHPDKNTDLQMRAKMGTPGTQMFVSIHVNATEASAALAGYGIETWWNPNHALSY
                     QLATVLQRDMVDVTGAFDRGIKGYRSLSVLRNSRIPAALVEIGFTSHPVDGQNLADNN
                     YLDRVAMGIASGIRDALMTGVSAAGKPAASAVGGGP"
     gene            complement(5770. .7146)
                     /gene="DR2395"
     CDS             complement(5770. .7146)
                     /gene="DR2395"
                     /note="similar to GP:3327262 percent identity: 64.01;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="Na+/H+ antiporter, putative"
                     /protein_id="AAF11941.1"
                     /db_xref="GI:6460212"
                     /translation="MWTWVCFRLEAPRVTDMLTAFALLLSVTAVLAFLNERFFRLPTT
                     VGVTLAGALSSMLLIALDALGLTGPRGWAAGLLETLDFTDFVLNGILSILLFAGALGL
                     DARQLLRQRFSILTLAVFSTLLSTFLIGFAAYGIFGLVGLNVPLLWSLLFGALISPTD
                     PVAVLDLLKRAAVPKRIETLIAGESLFNDGVGVVIFLVIAGMAGIGHSHGTGSVVGAL
                     ELFGREALGGMAFGALLGAIGFLMLREIEQHAVEVLITLALVVGGYVAASALGVSGPL
                     AMVVAGLVISAGRDVAFGEETREHIESFWETTDQVLNILLFAFIGLDVLLTETTGPQI
                     VAGLLLVVASLLARWISVALPFPLIRRHEDYGPYTVRLLTWGGLRGGIAISLVLGLPQ
                     TPYRTHLVTATYIIVLFSIAVQGLTIMPLVHKAVAASPDSAGPDSAGPGGGQEDGENA
                     GRQLSSET"
```

```
     gene            7192. .7788
                     /gene="DR2396"
     CDS             7192. .7788
                     /gene="DR2396"
                     /note="similar to GB:AL123456 percent identity: 52.33;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional repressor, TetR family"
                     /protein_id="AAF11942.1"
                     /db_xref="GI:6460213"
                     /translation="MPEPRRRLSAEVRREQILEVASQLFIERGFENVKMADIAEHLQT
                     SRPNIYTYYPSTEAILDILLERRMGDWLQEVDRRVADNSLFPPDDMLRALLEHRELLL
                     LLYSGGGPRFQSRRQRVLDSLDQHTYNYLPSNVTEANPDLILVQRTLVLSTAHELLLH
                     PEYDREQIGQLLNIMLVSAFKERVPDIDEQIQAKHSRA"
     gene            complement(7839. .8645)
                     /gene="DR2397"
     CDS             complement(7839. .8645)
                     /gene="DR2397"
                     /note="similar to GB:U00096 PID:1789101 PID:882637 percent
                     identity: 61.28; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="survival protein SurE"
                     /protein_id="AAF11943.1"
                     /db_xref="GI:6460214"
                     /translation="MTVPDPARPAASPSARPRVLVANDDGIFAPGIKALGLALSEWAD
                     VVVVAPDVEQSAVGHGITIRRPLRFKHTAAAGFGDIPAYRVDGTPADCVVLGVHLLGR
                     PDLVVSGINIGPNLGEDLTHSGTVAAAIEGLTLGLPSIAFSQFANEAGEYDFGPSAAY
                     ASRLAREVCCRGLPPRVLLNVNFPRVSPRGVRVTEVGLHRWEDSVVTRQDPEGRDYHW
                     VAGVSTAHDGHDEQTDYGAVQAGFISVSPVRLDLTARDLIGELTQALPPL"
     gene            8712. .9314
                     /gene="DR2398"
     CDS             8712. .9314
                     /gene="DR2398"
                     /note="similar to GP:2815343 percent identity: 58.15;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF11944.1"
                     /db_xref="GI:6460215"
                     /translation="MTVIRTLEDLESLFGEVGAPSLQKETDVLHSVYRQWLEAAPFAV
                     LATSGPGGLDASPRGDPAPLVRVTDERTLLLPERRGNNRIDSLRNLLSDPRVGLLFFI
                     PGVNETLRVNGTARILTTPDLLASFAQDGKLPKCVLEITVDTVFFQCGRAMLRSVLWT
                     GPSAAEVPTPGAMLAALTRGEVDGADYDAALPQRQRDTLY"
     gene            complement(9345. .10322)
                     /gene="DR2399"
     CDS             complement(9345. .10322)
                     /gene="DR2399"
                     /note="similar to GB:AL123456 percent identity: 54.93;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF11945.1"
                     /db_xref="GI:6460216"
                     /translation="MSAPLPCGALYPEPHDSCQGPRQWSGTELGWADGGRVGGRARPH
                     RRGTRPLAGRFSAQPAGAGRRAGAGPLDPRRAVPRAHLRDGAQFAHPEQPDEFTERVS
                     IFAFPPRTDSPHGAVLTHSRSGTRALGTVWALVGRETVNTAQEVTYELLDQTADGFFL
                     AADALEAQVDALEEQIFRSPRFNPVEPVFDLKHLLGQARRLTTDAREASALLGRQAST
                     AAEQVRYRDVQDSFTRVGSRLDGLRESLSSLLDLHLALQGQRMNEVMRTLTSVSVVFL
                     PLTFLAGVWGMKFQHMPELATRYGYAFAWLTFLLIGGLLTYWFKRRGWW"
BASE COUNT     1771 a   3750 c   3474 g   1519 t
ORIGIN


  Query Match             2.7%;  Score 36.8;  DB 1;  Length 10514;
  Best Local Similarity   47.1%;  Pred. No. 2.1e+02;
  Matches  113;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy   1018 gcgccgcagccggacaaatactccatcacgcgcaccactctcggccatttcctaaaaaac 1077
          |||| || |  || ||     | | | | ||    |    ||| ||  ||   |       |
Db   6450 GCGCTGCTGGGGGCCATCGGGTTCCTGATGCTGCGCGAAATCGAGCAGCACGCCGTCGAG 6391

Qy   1078 aaactcttcaagttcacgacagccgtcaacggcggcgaccgcgccatggtaccgatcggc 1137
           ||| |||   || ||   | ||||  |||   ||     |||  || || |  || |
Db   6390 GTGCTCATCACCCTCGCGCTGGTCGTCGGCGGGTACGTGGCGGCCTCGGCACTGGGCGTC 6331

Qy   1138 acttatgagcgcgtaatgccgttggacatcctgcctaccttgcttttgcgcgatttaatc 1197
          |       ||  |  |||  | | | |   |||   | || |     ||||||  |   |
Db   6330 AGCGGCCCGCTGGCGATGGTGGTCGCCGGTCTGGTCATCTCGGCGGGGCGCGACGTGGCC 6271

Qy   1198 gtcggcgataccgacagcgcgcaggctttgggttgcttggaattggacgaagaagacctc 1257
           |||| ||      | |||  ||      | ||| || ||||  |  |||  | |  |||
Db   6270 TTCGGGGAAGAAACCCGCGAACACATCGAGAGTTTCTGGGAAACGACCGATCAGGTGCTC 6211


RESULT  45
M23555
LOCUS       M23555                 3855 bp    DNA     linear   UNA 04-AUG-1993
DEFINITION  Figure 2. Nucleotide sequence of the 3,854-bp HindIII fragment
            encompassing the hrp locus and the deduced amino acid sequences of
            putative polypeptides encoded on the sense strand.
ACCESSION   M23555
VERSION     M23555.1  GI:340948
KEYWORDS    .
SOURCE      unclassified.
  ORGANISM  unclassified
            Unclassified.
REFERENCE   1  (bases 1 to 3855)
  AUTHORS   Mukhopadhyay,P., Williams,J. and Mills,D.
  TITLE     Molecular Analysis of a Pathogenicity Locus in Pseudomonas syringae
            pv. syringae
  JOURNAL   J. Bacteriol. 170, 5479-5488 (1988)
  MEDLINE   89053865
FEATURES             Location/Qualifiers
     source          1. .3855
                     /organism="unclassified"
                     /db_xref="taxon:12908"
BASE COUNT      741 a   1178 c   1181 g    755 t
ORIGIN


  Query Match             2.7%;  Score 36.4;  DB 13;  Length 3855;
  Best Local Similarity   45.9%;  Pred. No. 2.5e+02;
  Matches  124;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy    273 gattgccgttgaaggcaacgacgaaatcgagttcgaacgctacgtacctgaagcgctggc 332
          | |  ||  ||| | ||||| |||  | | |    |   |          |  || |  |
Db    615 GGTGACCACTGACGACAACGTCGAGCTGGTGGAAAACAACCTGCGCTACAACCCGGTCAC 674

Qy    333 aaaattgagcagcgaaaaagtgcgccgcaacctgattcaatcaggcttatggactgcgct 392
           |||     | ||   |   ||||  |||  | |   | ||  ||     |||  |   |
Db    675 CAAAGGTTACCGCCTGACCCTGCGTGTCAAGGTCAAGGATTCCAGCAAGCCGACCGAAAT 734

Qy    393 tcgcacccgtccgttcagcaaaatccctgccgtagatgccgagccgttcgccatcttcgt 452
           ||| ||   | |||  |  ||||||||||| |   | | ||      ||  |  ||||
Db    735 GCGCGCCTACCTGTTGCGTGAAATCCCTGCCGAACCGGGCAAGGAACCTGCGCTGCTCGT 794

Qy    453 caatgcgatggacaccaatccgctggctgccgaccctacggtcatcatcaaagaagccgc 512
            ||  |  | |    |   |  ||||||||       |    ||     | || | |
Db    795 GGCTGACAAAGCCGAAGAGAAGAAGGCTGCCGCGAAGGAAGCTGCCAAGCCGGCAGTCTC 854

Qy    513 cgaagacttcaaacgcggcctgttggtatt 542
          | | || | |     || || | | | | |
Db    855 CAAGGAGTCCGCCAACGACCAGGTAGAAAT 884


Search completed: June 30, 2002, 07:52:06
Job time: 52544 sec
```